PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:
C12N 1/21, 5/10, 15/11, 15/12, 15/63,
A61K 38/16, 38/17, C07K 14/00, 14/435,
16/00. G01N 33/50

(11) International Publication Number:

WO 00/06698

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number:

PCT/US99/17130

A1

(22) International Filing Date:

29 July 1999 (29.07.99)

(30) Priority Data:

60/094,657	30 July 1998 (30.07.98)	US
60/095,486	5 August 1998 (05.08.98)	US
60/095,455	6 August 1998 (06.08.98)	US
60/095,454	6 August 1998 (06.08.98)	US
60/096,319	12 August 1998 (12.08.98)	US

(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KOMATSOULIS, George, A. [US/US]; 9518 Garwood Street, Silver Spring, MD 20901 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). DUAN, Roxanne [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). MOORE, Paul, A. [GB/US]; 19005 Leatherbark Drive, Germantown, MD 20874 (US). SHI, Yanggu [CN/US]; Apartment 102,

437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David [US/US]; 3142 Quesada Street, N.W., Washington, DC 20015 (US). WEI, Ying-Fei [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; Apartment 115, 410 Van Dyke Street, St. Paul, MN 55119-4321 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). MUCENSKI, Michael [US/US]; 3265 Mandale Drive, Cincinnati, OH 45239 (US).

(74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report. With an indication in relation to deposited biological

material furnished under Rule 13bis separately from the description.

(54) Title: 98 HUMAN SECRETED PROTEINS

(57) Abstract

The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	0				
AM	Armenia		Spain	LS	Lesotho	SI	Slovenia
AT	Austria	FI	Finland	LT	Lithuania	SK	Slovakia
AU	·	FR	France	LU	Luxembourg	SN	Senegal
AZ	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
	Azerbaijan	GB	United Kingdom	MC	Моласо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	
BB	Barbados	GH	Ghana	MG	Madagascar	-	Togo
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TJ	Tajikistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TM	Turkmenistan
BG	Bulgaria	HU	Hungary	ML	Mali	TR	Turkey
BJ	Benin	IE	Ireland	MN		TT	Trinidad and Tobago
BR	Brazil	IL	Israel	MR	Mongolia Mauritania	UA	G Uganda
BY	Belarus	IS	Iceland	MW		UG	
CA	Canada	IT	Italy		Malawi	US	United States of America
CF	Central African Republic	JP	Japan	MX	Mexico	UZ	Uzbekistan
CG	Сопдо	KE	Kenya	NE	Niger	VN	Viet Nam
CH	Switzerland	KG	Kyrgyzstan	NL	Netherlands	YU	Yugoslavia
CI	Côte d'Ivoire	KP		NO	Norway	ZW	Zimbabwe
CM	Cameroon	N.	Democratic People's	NZ	New Zealand		
CN	China	KR	Republic of Korea	PL	Poland		
CU	Cuba		Republic of Korea	PT	Portugal		
cz	Czech Republic	KZ	Kazakstan	RO	Romania		
DE	Germany	LC	Saint Lucia	RU	Russian Federation		
DK		LI	Liechtenstein	SD	Sudan		
EE	Denmark	LK	Sri Lanka	SE	Sweden		
C.E.	Estonia	LR	Liberia	SG	Singapore		

WO 00/06698 PCT/US99/17130

98 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

5

10

15

20

25

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of

the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

5

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

15

20

25

30

10

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

10

15

20

25

30

In specific embodiments, the polynucleotides of the invention are less than 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, or 7.5 kb in length. In a further embodiment, polynucleotides of the invention comprise at least 15 contiguous nucleotides of the coding sequence, but do not comprise all or a portion of any intron. In another embodiment, the nucleic acid comprising the coding sequence does not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene in the genome).

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's

10

15

20

25

30

solution. 10% dextran sulfate, and 20 μ g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA

15

20

25

30

that may be single-stranded or, more typically, double-stranded or a mixture of single-and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

15

25

30

(See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

20 Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene is a human glycoprotein-associated amino acid transporter (See, e.g., Genbank Accession No. emblCAA10198.11 (AJ130718); all references available through this accession are hereby incorporated by reference herein). Amino acid transport across cellular membranes is mediated by multiple transporters with overlapping specificities. The transport system L, which mediates Na+-independent exchange of large neutral amino acids, consists of a novel amino acid permease-related protein (LAT1 or AmAT-L-lc) which for surface expression and function requires formation of disulfide-linked heterodimers with the glycosylated heavy chain of the h4F2/CD98 surface antigen. h4F2hc also associates with other mammalian light chains, e.g. y+LAT1 from mouse and human which are

10

15

20

25

30

approximately 48% identical with LAT1 and thus belong to the same family of glycoprotein-associated amino acid transporters.

The novel heterodimers form exchangers which mediate the cellular efflux of cationic amino acids and the Na+-dependent uptake of large neutral amino acids. These transport characteristics and kinetic and pharmacological fingerprints identify them as y+L-type transport systems. mRNA encoding my+LAT1 is detectable in most adult tissues and expressed at high levels in kidney cortex and intestine. This indicates that the y+LAT1-4F2hc heterodimer, besides participating in amino acid uptake/secretion in many cell types, is the basolateral amino acid exchanger involved in transepithelial reabsorption of cationic amino acids; hence, its defect might be the cause of the human genetic disease lysinuric protein intolerance.

The gene encoding the disclosed cDNA is believed to reside on chromosome 14. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 14.

Preferred polypeptides comprise the following amino acid sequence:

LALYSALFSYSGWDTLN (SEQ ID NO: 237), VTEEIKNPERNLPL (SEQ ID NO: 238), IGISMPIVT (SEQ ID NO: 239), IYILTNVAYYTVL (SEQ ID NO: 240), SDA VAVTFADQ (SEQ ID NO: 241), VALSCFGGLNASI (SEQ ID NO: 242), SRLFFV GSREGHLPD (SEQ ID NO: 243), SFSYWFFVGLS (SEQ ID NO: 244), VGQLYLR WKEP (SEQ ID NO: 245), RPRPLKLSVFFPIVFC (SEQ ID NO: 246), DTINSLIGI (SEQ ID NO: 247), LLAAACICLLTFINCAYVKWGTLVQDIFTYAKVLALIAVI VAGIVRLGQGASTHFENSFEGSSFAVGDIALALYSALFSYSGWDTLNYVTEEI KNPERNLPLSIGISMPIVTIIYILTNVAYYTVLDMRDILASDAVAVTFADQIFGIF NWIIPLSVALSCFGGLNASIVAASRLFFVGSREGHLPDAICMIHVERFTPVPSLL FNGIMALIYLCVEDIFQLINYYSFSYWFFVGLSIVGQLYLRWKEPDRPRPLKLS VFFPIVFCLCTIFLVAVPLYSDTINSLIGIAIALSGLPFYFLIIRVPEHKRPLYLRRI VGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN (SEQ ID NO: 249) and/or ATALPPKIVGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN (SEQ ID NO: 248). Polynucleotides encoding these polypeptides are also provided.

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 monocyte cells to calcium. Thus, it is likely that the product of this gene is involved in a signal

10

15

20

transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both THP-1 monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocytes.

This gene is expressed primarily in endothelial cells and brain, and, to a lesser extent, in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the neural or gastrointesinal systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulation system or central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 124 as residues: Glu-102 to Asn-110, Arg-256 to Leu-266, Pro-316 to Trp-328, Pro-331 to Arg-336, Met-350 to Gly-358. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain combined with its homology to a amino acid transporter and biological activity of increasing ion flux in monocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,

10

15

20

25

30

Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1550 of SEQ ID NO:11, b is an integer of 15 to 1564, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The gene encoding the disclosed cDNA is believed to reside on the X chromosome. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for the X chromosome.

Preferred polypeptides of the invention comprise the following amino acid sequence:

10

15

20

AARGSGVRDPLEEAVCPFSDLQLHAGRTTALFKAVRQGHLSLQRLLLSFVCL CPAPRGGAYRGRQASLSCGGLHPVRASRLLCLPKQAWAMAGAPPPVSLPPCS LISDCCASNQRDSVG (SEQ ID NO: 250). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in cord blood cells, and, to a lesser extent, in frontal lobe of the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, reproductive, hematopoietic or neural disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 125 as residues: His-56 to Gln-65, Leu-80 to Ile-85. Polynucleotides encoding said polypeptides are also provided.

25 polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

10

15

20

25

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1743 of SEQ ID NO:12, b is an integer of 15 to 1757, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in human T cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

10

15

20

25

30

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T cell lymphoma, immunodeficiencies, in addition to other immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 126 as residues: Met-1 to Phe-10. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human T cell lymphomas indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

10

15

20

25

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1359 of SEQ ID NO:13, b is an integer of 15 to 1373, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

The protein product of this clone shares sequence homology with the Cteminus of a human N-acetylglucosamine-phosphate mutase (See, e.g., Genbank Accession No. gblAAC72409.11 (AF102265); all references available through this accession are hereby incorporated by reference herein.) Hofmann, et al. (Eur. J. Biochem. 221:741-747 (1994)) studied the N-acetylglucosamine-phosphate mutase of 30 Saccharomyces cerevisiae and showed it to be essential for viability. A S. cerevisiae agm1 deletion mutant progressed through only approximately five cell cycles to form a 'string' of undivided cells with an abnormal cell morphology resembling

glucosamine auxotrophic mutants. Expression of the AGM1 gene on a multi-copy plasmid led to a significantly increased N-acetylglucosamine-phosphate mutase activity. Unlike over-expression of the S. cerevisiae AGM1 gene in a phosphoglucomutase (pgm1 delta/pgm2 delta) double deletion mutant which could restore phosphoglucomutase activity, over-expression of the PGM2 gene encoding the major isoenzyme of phosphoglucomutase did not increase N-acetylglucosamine-phosphate-mutase activity and did not restore growth of agm1 deletion mutant cells. These observations indicate that the different hexosephosphate mutases of S. cerevisiae have partially overlapping substrate specificities but, nevertheless, distinct physiological functions. The human N-acetylglucosamine-phosphate-mutase is expected to share at least some biological activities with the Agm1 protein.

10

Preferred polypeptide fragments of the invention comprise the following amino acid sequences: LSKAFLDSPNRLLAVEMNTDHLRLTVPNGIGALKLRXM EHYFSQGLSVQLFNDGSKGKLNHLCGADFVKSHQKPPQGMEIKSNERCCSFD 15 GDADRIVYYYHDADGHFHLIDGDKIATLISSFLKELLVEIGESLNIGVVQTAYA NGSSTRYLEEVMKVPVYCTKTGVKHLHHKAQEFDIGVYFEANGHGTALFST AVEMKIKQSAEQLEDKKRKAAKMLENIIDLFNQAAGDAISDMLVIEAILALK GLTVQQWDALYTDLPNRQLKVQVADRRVISTTXAERQAVTPPGLOEAINDL VKKYKLSRAFVRPSGTEDVVRVYAEADSQESADHLAHEVSLAVFOLAGGIGE 20 RPQPGF (SEQ ID NO: 251), LSKAFLDSPNRLLAVEMNTDHLRLTV (SEO ID NO: 252), PNGIGALKLRXMEHYFSQGLSVQLFNDG (SEQ ID NO: 253), SKGKL NHLCGADFVKSHQKPPQGMEIKS (SEQ ID NO: 254), NERCCSFDGDADRIV YYYHDADGHFHLI (SEQ ID NO: 255), DGDKIATLISSFLKELLVEIGESLNIGV (SEQ ID NO: 256), VQTAYANGSSTRYLEEVMKVPVYCTKTG (SEQ ID NO: 25 257), VKHLHHKAQEFDIGVYFEANGHGTALFS (SEQ ID NO: 258), TAVEMK IKQSAEQLEDKKRKAAKMLENI (SEQ ID NO: 259), IDLFNQAAGDAISDM LVIEAILALKGLT (SEQ ID NO: 260), VQQWDALYTDLPNRQLKVQVADRR VIST (SEQ ID NO: 261), TXAERQAVTPPGLQEAINDLVKKYKLSR (SEQ ID NO: 262), AFVRPSGTEDVVRVYAEADSQESA (SEQ ID NO: 263), and/or DH 30 LAHEVSLAVFQLAGGIGERPQPGF (SEQ ID NO: 264). Polynucleotides encoding these polypeptides are also provided.

10

15

20

25

30

The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in fetal brain, and, to a lesser extent, in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental disorders, particularly of the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central and peripheral nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 127 as residues: Asn-36 to Lys-42, Lys-53 to Gln-60, Ile-64 to Ala-77, Ala-128 to Tyr-135, Lys-184 to Ala-199, Leu-245 to Leu-250. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of N-acetylglucosamine-phosphate mutase in fetal brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital

15

20

30

malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3726 of SEQ ID NO:14, b is an integer of 15 to 3740, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in human stomach and stomach tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the gastrointestinal system, particularly cancer or ulcers of stomach tissue. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the

15

20

25

tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of the stomach indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment and intervention of these tumors, in addition to other tumors where expression has been indicated. Additionally, the protein product of this gene may play a role in the normal function of the stomach and/or digestive system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1182 of SEQ ID NO:15, b is an integer of 15 to 1196, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 6

Preferred polypeptides of the invention comprise the following amino acid sequences:

20

25

30

FEIALPRESNITVLIKLGTPTLLAKPCYIVISKRHITMLSIKSGERIVFTFSCQSPE
NHFVIEIQKNIDCMSGPCPFGEVQLQPSTSLLPTLNRTFIWDVKAHKSIGLELQ
FSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTFCSNGTVSRIKM (SEQ ID
NO: 266), and/or GTRAAPGLGAWGRRSPPSFSPPRPRRPGVMAGLNCGVSIAL

5 LGVLLLGAARLPRGAEAFEIALPRESNITVLIKLGTPTLLAKPCYIVISKRHITM
LSIKSGERIVFTFSCQSPENHFVIEIQKNIDCMSGPCPFGEVQLQPSTSLLPTLNR
TFIWDVKAHKSIGLELQFSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTFC
SNGTVSRIKMQEGVKMALHLPWFHPRNVSGFSIANRSSIKRLCIIESVFEGEGS
ATLMSANYPEGFPEDELMTWQFVVPAHLRASVSFLNFNLSNCERKEERVEYY

10 IPGSTTNPEVFKLEDKQPGNMAGNFNLSLQGCDQDAQSPGILRLQFQVLVQH
PQNESNKIYVVDLSNERAMSLTIEPRPVKQSRKFVPGCFVCLESRTCSSNLTLT
SGSKHKISFLCDDLTRLWMNVEKP (SEQ ID NO: 265). Polynucleotides encoding
these polypeptides are also provided.

This gene is expressed primarily in placenta, and to a lesser extent in, prostate and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, male and female infertility, and associated disorders of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in the prostate, placenta and ovary indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment, prevention, and/or diagnosis of male or female infertility, endocrine

disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia and prostate cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within placental tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2195 of SEQ ID NO:16, b is an integer of 15 to 2209, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where b is greater than or equal to a + 14.

25

30

10

15

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares homology with the human and rat HNK-1 sulfotransferase protein (See, e.g., Genbank Accession Nos. gblAAB88123.11 (AF022729) and gil2921306lgblAAC04707.11 (AF033827); all references available through these accessions are hereby incorporated herein by reference.) Ong E, et al. (J Biol Chem. 273(9):5190-5 (1998)) have characterized the human HNK-1 sulfotransferase, and show that it is involved in the synthesis of the HNK-1

carbohydrate epitope which is expressed on various adhesion molecules in the nervous system and on immune cells (e.g., natural killer cells) and is suggested to play a role in cell-cell and cell-substratum interactions. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with HNK-1 sulfotransferase proteins. Such activities are known in the art, some of which are described elsewhere herein, or in, for example, Bakker, et al., J Biol Chem. 272:29942-6 (1997), incorporated herein by reference. Based on sequence similarity between sulfotransferases, a consensus sequence for the active site was developed (Ong, et al., supra). The consensus pattern is as follows: xxRPDzzzz, where x represents hydrophobic amino acid residues and z represents any amino acid residue. Therefore,

Preferred polypeptides of the invention comprise the following amino acid sequences: FVRDPFVRL (SEQ ID NO: 267), FLFVRDPFVRLIS (SEQ ID NO: 268), FLFVRDPFVRLISAF (SEQ ID NO: 269), and/or YLHTSFSRPHTGPPLPTPG PDRDRELTADSDVDEFLDKFLSAGVKQSDLPRKETEQPPAPGSMEENVRGY DWSPRDARRSPDQGRQQAERRSVLRGFCANSSLAFPTKERAFDDIPNSELSHL IVDDRHGAIYCYVPKVACTNWKRVMIVLSGSLLHRGAPYRDPLRIPREHVH NASAHLTFNKFWRRYGKLSRHLMKVKLKKYTKFLFVRDPFVRLISAFRSK FELENEEFYRKFAVPMLRLYANHTSLPASAREAFRAGLKVSFANFIQYLLDPH TEKLAPFNEHWRQVYRLCHPCQIDYDFVGKLETLDEDAAQLLQLLQVDRQ LRFPPSYRNRTASSWEEDWFAKIPLAWRQQLYKLYEADFVLFGYPKPENLL RD (SEQ ID NO: 270). Polynucleotides encoding these polypeptides are also provided.

Further preferred are the sulfotransferase active site polypeptides listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the sulfotransferase active site polypeptides. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the sulfotransferase active site polypeptides, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to sulfotransferase proteins. The nucleotides sequence of this gene was found to be

15

20

25

 $\mathcal{F}_{\mathcal{F}}$.

homologous to the human hypoxanthine guanine phosphoribosyl transferase 2 cDNA which is know to be involved in the purine salvage pathway resulting in the maintainance of homeostatic levels of uric acid (See Genbank Accession No.T30127).

The gene encoding the disclosed cDNA is believed to reside on chromosome

7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

This gene is expressed to a very high level in HL-60 myelogenous leukemia cell lines, and to a lesser extent, in most cell types of the immune system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, myelopoiesis, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, metabolic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 130 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62, Lys-79 to Ala-93, Gly-95 to Asp-105, Ser-107 to Val-127, Gly-193 to Leu-200, Lys-218 to Ser-227, Lys-234 to Thr-239, Pro-366 to Asp-379, Pro-406 to Asp-414. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in HL-60 myelogenous leukemia cell lines and homology to HNK-1 sulfotransferase proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention and/or treatment of a variety of immune system disorders, including but not limited to, those involving the HNK-1 carbohydrate epitope, (e.g. HNK-1 as an auto-antigen

15

20

25

30

in peripheral demyelinative neuropathy). Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, expression of this gene product in tonsils indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the homology to a conserved purine metabolism protein may suggest that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sach's Disease, phenylkenonuria, galactosemia, porphyrias, Hurler's syndrome, and various urogenital disorders related to metabolic conditions, particularly Lesch-Nyhan syndrome. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15

20

25

30

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1760 of SEQ ID NO:17, b is an integer of 15 to 1774, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:17, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS) promoter element. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway, a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: KLVRLQVPVRNSRVDPRVRSKIGSRRWMLQLI MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVPYLYFDMGAAVLCA SFMSFGVKRRWFALGAALQLAISTYAAYIGGYVHYGDWLKVRMYSRTVAII GGFLVLASGAGELYRRKPRSRSLQSTGQVFLGIYLICVAYSLQHSKEDRLA YLNHLPGGELMIQLFFVLYGILALAFLSGYYVTLAAQILAVLLPPVMLLIDG NVAYWHNTRRVEFWNQMKLLGESVGIFGTAVILATDG (SEQ ID NO: 271).

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVP YLYFDMGAAVLCASFMSFGVKRRWFALGAALQLAISTYAAYIGGYVHYGD

10

15

20

25

30

WLKVRMYSRTVAIIGGFLVLASGAGELYRRKPRSRSLQSTGQVFLGIYLICVA YSLQHSKEDRLAYLNHLPGGELMIQLFFVLYGILAPGLSVRLLRDPRCPDPGC TAAPCHAAH (SEQ ID NO: 272). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 17. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 17.

This gene is expressed primarily in endometrial tumors, and to a lesser extent, in T-cells, pituitary and to a certain extent in most cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, female reproductive, immune, or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and/or immune systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 131 as residues: Ala-27 to Asp-34, Tyr-116 to Leu-125. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution predominantly in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of a range of immune and/or reproductive disorders including endometriosis, endometritis, and endometrioma. Similarly, the tissue distribution in T-cells and the ability of supernatants expressing this gene to stimulate the GAS promoter element in T-cells indicates polynucleotides and

10

15

20

25

30

polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in pituitary indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of the Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-,

10

15

20

25

30

hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-,hypoparathyroidism), hypothallamus, and testes.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1660 of SEQ ID NO:18, b is an integer of 15 to 1674, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of the myeloid leukemia cell line AML-193 to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of myeloid leukemia cells, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating myeloid leukemia cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: SNEILLSFPQNYYIQWLNGSLIHGLWNLASLFS NLCLFVLMPFAFFFLESEGFAGLKKGIRARILETLVMLLLLALLILGIVWVAS ALIDNDAASMESLYDLWEFYLPYLYSCISLMGCLLLLLCTPVGLSRMFTVMG HLLVKPTILEDLDEQIYIITLEEEALQRRLNGLSSSVEYNIMELEQELENVKTL KTKLERRKKASAWERNLVYPAVMVLLLIETSISVLLVACNILCLLVDETAM PKGTRGPGIGNASLSTFGFVGAALEIILIFYLMVSSVVGFYSLRFFGNFTPKKD DTTMTKIIGNCVSILVLSSALPVMSRTLGITRFDLLGDFGRFNWLGNFYIVLS YNLLFAIVTTLCLVRKFTSAVREELFKALGLHKLHLPNTSRDSETAKPSVNGH

10

15

20

25

30

QKAL (SEQ ID NO: 273). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in fetal heart, and to a lesser extent, in colon and the adult pulmonary system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, heart, lung and digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, pulmonary and digestive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 132 as residues: Glu-67 to Asn-74, Glu-88 to Asn-93, Lys-95 to Ala-107, Ala-147 to Arg-153, Phe-197 to Thr-205, Pro-292 to His-308. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal heart, colon and pulmonary tissues and the biological activity in increasing the permeability of the plasma membrane of the myeloid leukemia cell line AML-193 to calcium, likely indicating that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of myeloid leukemia cells, indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention, and/or detection of a range of disorders including a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, myocardial infarction, myocarditis, ischemia, thrombosis, coronary artery disease,

10

15

arteriosclerosis, and/or atherosclerosis; pulmonary edema and embolism, bronchitis and/or cystic fibrosis; Crohn's Disease and/or colon cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2004 of SEQ ID NO:19, b is an integer of 15 to 2018, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where b is greater than or equal to a + 14.

25

30

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

The protein product of this clone shares sequence homology with the human MaxiK channel beta 2 subunit (See, Genbank Accession No. gblAAD23380.1IAF099137_1 (AF099137); all references available through this accession are hereby incorporated herein by reference), which is believed to be a modulatory subunit of the voltage and Ca2+ activated K+ (MaxiK) channel. Additionally, this protein shares homology to the human calcium-activated potassium

20

30

channel beta subunit, which, when combined with its corresponding alpha subunit and modulating peptide, are believed to be useful in treating asthma, angina, hypertension, incontinence, migraine, irritable bowel syndrome (IBS). The subsequent heteromultimer that forms upon combining the alpha, beta, and modulator subunits are also thought to be useful in preventing premature labour, preventing and treating cerebral ischemia, inducing pain modulation and decreasing neurogenic inflammation in a patient (See GeneSeq Accession No. R85306).

Preferred polypeptides comprise the soluble domain which consists of the following amino acid sequence: RSYMQSVWTEESQCTLLNASITETFNCSFSCGP DCWKLSQYPCLQVYVNLTSSGEKLLLYHTEETIKINQKCSYIPKCGKNFEESM SLVNVVMENFRKYQHFSCYSDPEGNQKSVILTKLYSSNVLFHSLFWPTCMMA GGVAIVAMVKLTQYLSLLCERIQRINR (SEQ ID NO: 274). Polynucleotides encoding these polypeptides are also provided. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with modulatory subunits of voltage and Ca2+ activated K+ channel proteins. Such activities are known in the art, some of which are described in Wallner, et al., PNAS 96:4137-4142 (1999), incorporated herein by reference.

This gene is expressed primarily in adrenal gland tumor, and to a lesser extent, in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endocrine and immune disorders, particularly Hodgkin's Lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and/or endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression

10

15

20

25

30

level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 133 as residues: Trp-25 to Gln-30, Pro-50 to Gln-57, Pro-93 to Glu-101, Arg-114 to Cys-121, Ser-123 to Gln-129, Ile-177 to Arg-182. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in adrenal gland tumor and it's identification as the modulatory subunit of the voltage and Ca2+ activated K+ (MaxiK) channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-,hypoparathyroidism), hypothallamus, and testes. Alternatively, expression in proliferative immune tissues combined with its homology to a novel human K channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in Hodgkin's lymphoma indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

15

20

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2084 of SEQ ID NO:20, b is an integer of 15 to 2098, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

The translation product of this gene shares homology with collagen and collagen like proteins (See, e.g., Genbank Accession Nos. gil2920535lgblAAC39658.1l (AF018081) and gil2384942lgblAAB69961.1l (AF022985); all references available through these accession numbers are hereby incorporated by reference herein). Additionally, it has been determined that this gene has homology to the human Kruppel related zinc finger protein (HTF10) which is known to be important as a transcription factor, particularly in development (See Genebank Accession No.L11672).

15

20

25

30

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AFAHLQLGPMWKLWRAEEGAAALGGALFLLL FALGVRQLLKQRRPMGFPPGPPGLPFIGNIYSLAASSELPHVYMRKQSQVYG EVQPRRAPGREGRQAGPGWPGPSWLDLWPPLGRLVGTSPCAGCPLRDTRFPG LEGRS PRRRAPLQGEPRPCR (SEQ ID NO: 275). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in human erythroleukemia cell line (HEL), serum induced smooth muscle, and to a lesser extent in human 8 week whole embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia, musculoskeletal, or developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system and muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, musculoskeletal, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 134 as residues: Leu-30 to Gly-38, Arg-67 to Val-72, Val-76 to Ala-89, Pro-118 to Arg-123, Gly-129 to Ala-136, Leu-138 to Arg-146. Polynucleotides encoding said polypeptides are also provided.

fluid from an individual not having the disorder.

The tissue distribution in human erythroleukemia cell line (HEL), and serum induced smooth muscle, and the shared homology with collagen and collagen like proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for disorders of hematopoietic or muscular systems, such as leukemia and

15

25

30

muscular dystrophy. Additionally, the shared homology with collagen proteins would suggest that this protein may also be important in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.).

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating 20 infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of

supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15

20

25

30

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1732 of SEQ ID NO:21, b is an integer of 15 to 1746, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRVRIGLTLLLCAVLLSLASASSDEEGSQD ESLGFQDYFDIR (SEQ ID NO: 276). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 8. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 8.

This gene is expressed primarily in dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 135 as residues: Ser-22 to Ser-41. Glu-43 to Thr-50,

15

20

25

30

Ser-63 to Leu-68, Ser-71 to Gly-84, Ser-96 to Gly-114. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation

15

20

25

30

modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:22 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2862 of SEQ ID NO:22, b is an integer of 15 to 2876, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:22, and where b is greater than or equal to a + 14.

utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

A preferred polypeptide variant of the invention comprises the following amino acid sequence: MARGSLRRLLRLLVLGLWLALLRSVAGEQAPGTAPC SRGSSWSADLDKCMDCSTSCPLPA ALAHPWGRSEPDLRAGAAFWLFGLE

4.

5

20

25

30

TMPQE REVHHPHRGDRRRGLPSCGADPVTMCPLPAGARPLIIHSSILEPVSAS QTRREPSSSNHK GGGGR (SEQ ID NO: 277). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 16. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 16.

This gene is expressed primarily in tumor growth factor or lipopolysaccharide treated bone marrow stroma, epithelioid sarcoma, umbilical vein endothelial cells, and to a lesser extent, in other tissues.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoiesis or immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological 15 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic, integumentary, or immune systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic, immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 136 as residues: Pro-35 to Trp-42, Pro-65 to Asp-72. Thr-86 to Phe-93, Ile-97 to Glu-103. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in tumor growth factor or lipopolysaccharide treated bone marrow stroma, epithelioid sarcoma, and umbilical vein endothelial cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious

10

15

20

25

30

disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for

15

25

30

control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease): as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1038 of SEQ ID NO:23, b is an integer of 15 to 1052, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where b is greater than or equal to a + 14.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 14

The translation product of this gene shares sequence homology with chromaffin granule amine transporter protein which is thought to be important in vesicle membrane amine transport, particularly in the neural and endocrine tissue, and the human vesicular monoamine transporter hVMAT1 which is involved in the regulation of amine storage in cardiovascular, endocrine, and central nervous system function (See, Genbank Accession Nos. gil1314290 and gblAAC50472.11; all references available through these accession numbers are hereby incorporated by reference herein). Based on these sequence similarities, The translation product of this gene is expected to share at least some biological activities with amine transporter proteins. Such activities are known in the art, some of which are described in Erickson, et al., PNAS 93:5166-5171 (1996), and/or Liu, et al., Cell 70:539-551 (1992), which are both incorporated herein by reference.

15

20

25

30

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSFLDPTLSLFVLEKFNLPAGYVGLVFLGMAL SYAISSPLFGLLSDKRPPLRKWLLVFGNLITAGCYMLLGPVPILHIKSQLWLL VLILVVSGLSAGMSIIPTFPEILSCAHENGFEEGLSTLGLVSGLFSAMWSIGAF MGPTLGGFLYEKIGFEWAAAIQGLWALISGLAMGLFYLLEYSRRKRSKSQNIL STEEERTTLLPNET (SEQ ID NO: 278). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in colon cancer, osteoclastoma, and T-cell lymphoma, and to a lesser extent in many tumor or proliferative tissues such as endometrial tumor, chondrosarcoma, induced umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases resulting from disorders in small molecule transport (i.e., signalling molecules) in afflicted tissues and organs, particularly of the endocrine and central nervous systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal, immune, and/or digestive systems and cancer expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

 $f_{i,j}(\xi)$

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 137 as residues: Ser-114 to Asn-123, Thr-127 to Thr-132. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in colon cancer, osteoclastoma, and T-cell lymphoma and homology to amine transporter family members indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders or diseases resulted from small molecule transport in afflicted tissues and organs, particularly that of colon, osteoclast or T-cells. The expression in cancer tissues, and shared homology with transporter proteins may also indicate its role in anti-cancer drug resistance. Additionally, the protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; or for identifying inhibitors or promoters of the transport of toxic molecules to vesicles, for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of

10

15

20

25

30

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1527 of SEQ ID NO:24, b is an integer of 15 to 1541, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with the human prolyl 4-hydroxylase alpha (II) subunit which is important in catalyzing the formation of 4-hydroxyproline in collagens which is essential for the folding of newly synthesised collagen polypeptide chains into triple-helical molecules (See Genbank Accession No. gblAAB71339.11; all references available through this accession are hereby incorporated herein by reference). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Prolyl 4-hydroxylase proteins. Such activities are known in the art, some of which are described in Annunen, et al., J. Biol. Chem. 272:17342-17348 (1997) which is incorporated herein by reference.

When tested against U937 myeloid and Jurkat T-cell cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates myeloid cells and T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence:

10

15

20

25

30

GTREARLRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSL
EASENIRALKDGYEKVEQDLPAFEDLEGAARALMRLQDVYMLNVKGLAR
GVFQRVTGSAITDLYSPKRLFSLTGDDCFQVGKVAYDMGDYYHAIPWLEEA
VSLFRGSYGEWKTEDEASLEDALDHLAFAYFRAGNVSCALSLSREFLLYSPD
NKRMARNVLKYERLLAESPNHVVAEAVIQRPNIPHLQTRDTYEGLCQTL
GSQPTLYQIPSLYCSYETNSNAYLLLQPIRKEVIHLEPYIALYHDFVSDSEAQ
KIRELAEPWLQRSVVASGEKQLQVEYRISKSAWLKDTVDLKLVTLNHRIAA
LTGLDVRPPYAEYLQVVNYGIGGHYEPHFDHATSPSSPLYRMKSGNRVATFM
IYLSSVEAGGATAFIYANLSVPVVRNAALFWWNLHRSGEGDSDTLHAGCP
VLVGDKWVANKWIHEYGQEFRRPCSSSPED (SEQ ID NO: 282). Additional,

Preferred polypeptides comprise the following amino acid sequence: GTREA RLRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSLEASENI RALKDGYEKVEQ DLPAFEDLEGAARAL (SEQ ID NO: 279), ALMRLQD (SEQ ID NO: 280), and/or VEAGGAT (SEQ ID NO: 281). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in lymph node breast cancer, colon carcinoma, and to a lesser extent in osteoblasts and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of connective and immune tissues, particularly autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissues in breast, colon, bone, and fat, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, connective, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 138 as residues: Ser-74 to Ala-84, Gln-156 to Tyr-161, Tyr-184 to Asn-189, Ser-218 to Ile-223, Pro-299 to Ser-308, His-359 to Thr-368, Tyr-390 to Asp-404. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in lymph node breast cancer and colon carcinoma and homology to prolyl 4-hydroxylase alpha (II) subunit indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of connective tissue disorders and diseases (e.g. arthritis, trauma, tendonitis, chrondomalacia and inflammation), as well as, in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias ie. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Alternatively, the tissue distribution within various tissue carcinomas and tumor tissues, and biological activity reflected by the binding and activation of the GAS promoter element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders.

Expression in cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2065 of SEQ ID NO:25, b is an

10

15

20

25

30

integer of 15 to 2079, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

In an additional embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: IQPSHAALLHCRSTFRKTECLDPW WVRRQLLGMAGIGGLQKMKAPHTGVLHLGSVWVFLGPFLLGVGYTLTFNPL SGCMSTVRWLNSNITANRTLSRSVCHVTPLHRSLSPHDGEYLRQMLLNSSSR AGEAGSWGY (SEQ ID NO: 283). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 20. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 20.

This gene is expressed primarily in fetal liver, and, to a lesser extent, in a variety of fetal and other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, liver disorders and cancers (e.g., hepatoblastoma, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hepatic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, bile, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 139 as residues: Ser-67 to Tyr-75. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various would-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1933 of SEQ ID NO:26, b is an integer of 15 to 1947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this gene shares sequence homology with human laminin B1 which is thought to be an important structural extracellular matrix component involved in cell migration and signalling, paricularly in stimulating epithelial cell growth and differentiation (See, Genbank Accession No gil186837).

WO 00/06698 PCT/US99/17130

Preferred polypeptides of the invention comprise the following amino acid sequences: CSSPPGRLPWCWTAPRTLGKHGSLISTLRLTAPLHLAWKMMLS RKALFVLLNTPVLFHALEGRLFSKLCHHHTIQRTLTVPKFRSS (SEQ ID NO: 284), RSPTSRVQLLKRQSCPCQRNDLNEEPQHFTHYAIYDFIVKGSCFCNG HADQCIPVHGFRPVKAPGTFHMVHGKCM (SEQ ID NO: 285), and/or HNTAG SHCQHCAPLYNDRPWEAADGKTGAPNECRTCKCNGHADTCHFDVNVWEAS GNRSGGVCDDCQHNTEGQYCQRCKPGFYRDLRRPFSAPDACKPCSCHPV GSAVLPANSVTFCDPSNGDCPCKPGVAGRRCDRCMVGYWGFGDYGCRP CDCAGSCDPITGDCISSHTDIDWYHEVPDFRPVHNKSEPAWEWEDAQGFSAL LHSGKCECKEQTLGNAKAFCGMKYSYVLKIKILSAHDKGTHVEVNVKIK KVLKSTKLKIFRGKANIISRIMDGQ RMHLSNPQSWFGIPCSRT (SEQ ID NO: 286). Polynucleotides encoding these polypeptides are also provided.

Included in this invention as preferred domains are Laminin-type EGF-like (LE) domain signatures, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Laminins are the major noncollagenous components of basement membranes that mediate cell adhesion, growth migration, and differentiation. They are composed of distinct but related alpha, beta and gamma chains. The three chains form a cross-shaped molecule that consist of a long arm and three short globular arms. The long arm consists of a coiled coil structure contributed by all three chains and cross-linked by interchain disulfide bonds. Beside different types of globular domains each subunit contains, in its first half, consecutive repeats of about 60 amino acids in length that include eight conserved cysteines. The tertiary structure of this domain is remotely similar in its N-terminal to that of the EGF-like module. It is known as a 'LE' or 'laminin-type EGF-like' domain. The number of copies of the LE domain in the different forms of laminins is highly variable; from 3 up to 22 copies have been found. A schematic representation of the topology of the four disulfide bonds in the LE domain is shown below.

10

15

20

25

30

'C': conserved cysteine involved in a disulfide bond

'a': conserved aromatic residue

'G': conserved glycine (lower case = less conserved)

's': region similar to the EGF-like domain

'*': position of the pattern

In mouse laminin gamma-I chain, the seventh LE domain has been shown to be the only one that binds with a high affinity to nidogen. The binding-sites are located on the surface within the loops C1-C3 and C5-C6. Long consecutive arrays of LE domains in laminins form rod-like elements of limited flexibility, which determine the spacing in the formation of laminin networks of basement membranes. We derived a signature pattern for the LE domain which covers the C-terminal half of the repeat starting with the fourth conserved cysteine. The consensus pattern is as follows: C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C [All C's are involved in disulfide bonds]

Preferred polypeptides of the invention comprise the following amino acid sequence: CDDCQHNTEGQYCQRCKPGFYRDLRRPFSAPDACKPC (SEQ ID NO: 287) and/or CPCKPGVAGRRCDRCMVGYWGFGDYGCRPCDCAGSC (SEQ ID NO: 288). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the laminin-type EGF-like domains listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence encoded by this gene. The additional contiguous amino acid residues is N-terminal or C- terminal to the laminin-type EGF-

15

20

25

30

like domain. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the laminin-type EGF-like domain, wherein the total N-and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domain is characteristic of a signature specific to Laminin proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Laminin proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in osteoblastic tissues and cell types, including osteoblasts, osteoblastomas and osteoclastomas. Expression is also abundant in vascular-pulmonary tissues such as lung, micro-vasculature, pulmonary, endoithelial and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and malignancies (particularly of osteoblastic tissues and rhabdomyosarcoma), as well as cardiovascular and respiratory or pulmonary disorders such as athsma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stoke, thrombosis hypertension, inflammation and wound healing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardio-respiratory system. and skeletal system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, osteoblast, cardiorespiratory, vascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph. pulmonary surfactant, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 140 as residues: Ser-28 to Cys-34, Thr-51 to Thr-58, Tyr-64 to Asn-81, Asp-111 to Lys-116, Asp-145 to Phe-160, Pro-203 to Glu-217. Polynucleotides encoding said polypeptides are also provided.

15

20

25

The tissue distribution in osteoblastic tissues and cell types and homology to laminin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and diagnosis of cardiovasular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stoke, angina, thrombosis hypertension, inflammation and wound healing. As a homolog of laminin, this gene product quite possibly has a role in cell adhesion, migration, proliferation, angiogenesis, chondrogenesis, wound healing and oncogenesis. An EST (Int J Cancer 1996 May 16;66(4):571-577) with an identical sequence to part of this contig was shown to be differentially expressed in human primary myoblasts and embryonal rhabdomyosarcoma and therefore might have an important role in the determination or maintenance of the normal phenotype, and thus. its loss is possibly involved in the progression of malignancies, particularly of skeletal muscle. Similarly, the homology to a laminin would suggest a role in the detection and treatment of disorders and conditions afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chrondomalacia and inflammation) in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

30

formula of a-b, where a is any integer between 1 to 3365 of SEQ ID NO:27, b is an integer of 15 to 3379, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The gene encoding the disclosed cDNA is believed to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: NISSQYCILKSLEMMISGLKLLVLFLKFAPENY CLSTETLQMPNRHLRLSKATCYLMKCLLPSYFE (SEQ ID NO: 289). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta, brain, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, behavioral, or nervous system disorders, such as: depression, schizophrenia, Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, dementia, paranoia, addictive behavior, epilepsy, transmissible spongiform encephalopathy (TSE), Creutzfeldt-Jakob disease (CJD). Other diseases and conditions related to expression in the placenta might include developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive disfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and reproductive systems. expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial

10

15

20

25

30

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 141 as residues: Ala-16 to Leu-22. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, expression in placenta would suggest a possible role in the treatment and diagnosis of developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive disfunction and pre-natal disorders.

Similarly, expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may

20

25

30

show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1992 of SEQ ID NO:28, b is an integer of 15 to 2006, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

The translation product of this gene shares sequence homology with the murine transforming protein (See, e.g., Genbank Accession No. gil53529lemblCAA36859.11; all references available through this accession are hereby incorporated by reference herein).

This gene is expressed primarily in activated and basal T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, auto-immunities, cancer, metastasis, wound healing, inflammation, anemias (leukemia) and other hematopoietic disorders. Similarly, polypeptides and antibodies directed to these

10

15

20

25

30

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in activated T-cells and the homology to a murine transforming protein indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of

30

various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate igands or receptors, to identify agents 5 that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human 10 immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating 15 infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy 20 procedures. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3056 of SEQ ID NO:29, b is an

20

25

integer of 15 to 3070, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

The translation product of this gene was shown to have homology to the Mus musculus ALG-2 protein, which is known to code for a Ca(2+)-binding protein required for T cell receptor-, Fas-, and glucocorticoid-induced cell death. ALG-2 mediate Ca(2+)-regulated signals along the death pathway and may play a role in the onset of Alzheimer's Disease (See e.g., Genbank Accession No.gil1213520; all references available through this accession are hereby incorporated by reference herein).

Preferred polypeptides comprise the following amino acid sequence: NWVPT CLCPSAPCSFHLLSRFKCLFSPQRLTDIFRRYDTDQDGWIQVSYEQYLSMVFS IV (SEQ ID NO: 291), and/or QRLTDIFRRYDTDQDGWIQVSYEQYLSMVFSIV (SEQ ID NO: 292). Polynucleotides encoding these polypeptides are also provided.

When tested against K562 cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element). Thus, it is likely that this gene activates immune or leukemia cells through the Jaks-STAT signal transduction pathway. ISRE is a promoter element found upstream in many genes which are involved in the Jaks-STAT pathway. The Jaks-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MFYKLTLILCELSVAGVTQAASQRPLQRLPRHICSQR XPPGRCLLKAXLQTTWXXPDKPI PRLSPPLXSDPKR (SEQ ID NO: 293). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome

5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

15

20

25

30

This gene is expressed primarily in placenta, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 143 as residues: Arg-24 to Arg-31, Ile-33 to Gly-41. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and/or diagnosis of developmental anomalies, fetal deficiencies, ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells combined with the observed ISRE activity, and homology to the apoptosis linked, ALG-2 indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

15

20

25

30

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:30 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2213 of SEQ ID NO:30, b is an integer of 15 to 2227, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:30, and where b is greater than or equal to a + 14.

10

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

The translation product of this gene was shown to have homology to the human histo-blood group A transferase (See, e.g., Genbank Accession No. gblAAD26573.1lAF134413_1 (AF134413); all references available through this accession are hereby incorporated by reference herein) which is known to represent one of the major allogeneic antigens in both erythrocytes and tissues of humans. Its been proposed that the A phenotype is associated with the glycosyltransferas that converts the H substance associated with the O phenotype to A through the addition of alpha1-3-N-acetylgalactosamine or alpha1-3-galactosyl residues to the H antigen Fuc-alpha1-2Gal- beta1-R. Therefore, the primary product of the histo-blood group A is its respective glycosyltransferase. Preferred polypeptides of the invention comprise the following amino acid sequence: TSSPVFSFCSMAVREPDHLQ RVSLPRYNVSASLQWLPCHRIVLQPWHMCAMWELGQVLFHPVAPREGAAPS PVSTLTWPSSCSHSESTMELELQF (SEQ ID NO: 294), LPCHRIV (SEQ ID NO: 296), SLQWLPCHRIVLQPW (SEQ ID NO: 297), and/or MAVREPDHLQRVSLPR (SEQ ID NO: 295). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in 12-week-old human embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, pre-natal disorders, hematopoietic disorders, or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic, lymph, developing, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

The tissue distribution in 12 week old embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental anomalies, fetal deficiencies, pre-natal disorders and cancer. Similarly, expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Alternatively, the tissue distribution and homology to human blood group A and B glycosyltransferase enzymes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:31 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically

15

20

25

30

excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1274 of SEQ ID NO:31, b is an integer of 15 to 1288, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:31, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

The translation product of this gene shares sequence homology with CD97 10 (EMR1), which is thought to be important in both adhesion and signaling processes early after leukocyte activation (See, e.g., Genbank Accession No. gil784994; all references available through this accession are hereby incorporated by reference herein). EMR1 belongs to a novel family of G-protein receptors that has recently been recognized on the basis of homologous primary amino acid sequences, comprises receptors to hormones of the secretin/vasoactive intestinal peptide/glucagon family, parathyroid hormone and parathyroid hormone-related peptides, growth hormonereleasing factor, corticotropin-releasing factor, and calcitonin. Proteins with seven transmembrane segments (7TM) define a superfamily of receptors (7TMreceptors) sharing the same topology: an extracellular N-terminus, three extramembranous loops on either side of the plasma membrane, and a cytoplasmic C-terminal tail. Upon ligand binding, cytoplasmic portions of the activated receptor interact with heterotrimeric G-coupled proteins to induce various second messengers, which subsequently activate various signal transduction pathways depending upon the specific G-coupled protein associated with the receptor. Preferred polypeptides of the invention comprise the following amino acid sequence: CFKRKPKREHCSCP ITYQSLGDILNASFFSKRKGMQEVKLNSYVVSGTIGLKEKISLSEPVFLTFRHN QPGDKRTKHICVYWEGSEGGRWSTEGCSHVHSNGSYTKCKCFHLSSFAVLV ALAPKEDPVLTVITQVGLTISLLCLFLAILTFLLCRPIQNTSTSLHLELSLCLFLA HLLFLTGINRTEPEVLCSIIAGLLHFLYLACFTWMLLEGLHLFLTVRNLKVAN YTSTGRFKKRFMYPVGYGIPAVIIAVSAIVGPQNYGTFTHCWLKLDKGFIWSF MGPVAVIILINLVFYFQVLWILRSKLSSLNKEVSTIQDTRVMTFKAISOLFILGC SWGLGFFMVEEVGKTIGSIIAYSFTIINTLQGVLLFVVHCLLNRQVRMEYKKW

10

15

30

FSGMRKGVETESTEMSRSTTQTKTEEVGKSSEIFHKGGTASSSAESTKQPQPQ VHLVSAAWLKMN (SEQ ID NO: 298), and/or FFWKENLRRNGSREDFARRATQ LIQSVELSIWNASFASPGKGQISEFDIVYETKRCNETRENAFLEAGNNTMDINC ADALKGNLRESTAVALSLINLLGIF SEQ ID NO: 299. Polynucleotides encoding these polypeptides are also provided.

Included in this invention as preferred domains are two EGF-like protein domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). First, a sequence of about forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown to be present in a large number of membrane-bound and extracellular, mostly animal proteins. Many of these proteins require calcium for their biological function and a calcium-binding site has been found to be located at the N-terminus of some EGF-like domains. Calcium-binding is crucial for numerous protein-protein interactions. We have used the N-terminal part of the EGF domain as a consensus pattern. It includes the negative N-terminus and the possible hydroxylation site. The consensus pattern is as follows:[DEQN].[DEQN]{2}C.{3,14}C.{3,7}C.[DN].{4}[FY].C [The four C's are involved in disulfide bonds].

Preferred polypeptides of the invention comprise the following amino acid sequence: DINECETGLAKCKYKAYCRNKVGGYIC (SEQ ID NO: 300).

Polynucleotides encoding these polypeptides are also provided. Secondly, post-translational hydroxylation of aspartic acid or asparagine to form erythro-beta-hydroxyaspartic acid or erythro-beta-hydroxyasparagine has been identified in a number of proteins with domains homologous to (EGF). Based on sequence comparisons of the EGF-homology region that contains hydroxylated Asp or Asn, a consensus sequence located in the N-terminal of EGF-like domains has been identified that seems to be required by the hydroxylase(s). The consensus sequence is as follows: C.[DN].{4}[FY].C.C.

Preferred polypeptides of the invention comprise the following amino acid sequence: CRNKVGGYICSC (SEQ ID NO: 301). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the calcium-binding EGF-like domain and aspartic acid and asparagine hydroxylation site listed above, and at least

10

15

20

25

5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene and the embodiments listed herein. The additional contiguous amino acid residues is N-terminal or C- terminal to one or both of the listed domains. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to one or both of the listed domains, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to EGF like proteins. Based on the sequence similarity and conserved domains, The translation product of this gene is expected to share at least some biological activities with EGF-like proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in eosinophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders or anemias and leukemias, immunodeficiencies, infection, lymphomas, auto-immunities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 145 as residues: Ser-22 to Ser-30, Pro-33 to Cys-48, Asp-50 to Lys-67, Pro-117 to Ser-130. Polynucleotides encoding said polypeptides are also provided.

15

20

25

30

The tissue distribution in eosinophils combined with its homology to a known human seven transmembrane domain protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders, particularly considering the fact that the majority of 7 transmembrane receptors are tightly associated with signal transduction pathways which are integral to the modulation of the cell cycle. As such, the protein product of this gene may play a role in the regulation of cellular division, where loss of regulation may result in proliferating cells and the onset of tumors or cancer. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Similarly, the tissue distribution and homology to CD97 indicates that the protein product of this gene might be a marker for differentiation and activation of eosinophils, and therefore is useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g., AIDS), immuno-supressive conditions (transplantation) and hematopoietic disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

25

related to SEQ ID NO:32 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3266 of SEQ ID NO:32, b is an integer of 15 to 3280, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:32, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

The translation product of this gene has been found to have homology to the rat neural F box protein NFB42, in addition to a conserved Caenorhabditis elegans C14B1.3 protein (See, e.g., Genbank Accession Nos. gil3851648lgblAAC97505.11 (AF098301) and gil558270; all references available through these accessions are 15 hereby incorporated by reference herein). Preferred polypeptides of the invention comprise the following amino acid sequence: ALCPHPHLILNVTVSPAPSCRHVK KVVASPSPSTTMIAMDAPHSKAALDSINELPENILLELFTHVPARQLLLNCRL VCSLWRDLIDLMTLWKRKCLREGFITKDWDQPVADWKIFYFLRSLHRNLLR NPCAEEDMFAWQIDFNGGDRWKVESLPGAHGTDFPDPKVKKYFVTSYEMCL 20 KSQLVDLVAEGYWEELLDTFRPDIVVKDWFAARADCGCTYQLKVQLASA DYFVLASFEPPPVTIQQWNNATWTEVSYTFSDYPRGVRYILFQHGGRDTQY WAGWYGPRVTNSSIVVSPKMTRNQASSEAQPGQKHGQEEAAQSPYRAVV QIF (SEQ ID NO: 302). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune cells, especially primary dendritic cells, and T cells, and to a lesser extent in a variety of other tissues including breast, keratinocytes, epididiymus (cauda), lung, multiple sclerosis, endometrial stromal cells, IL4 induced umbilical vein endothelial cells, fetal kidney, fetal dura mater, rejected kidney, and osteoblasts.

15

20

25

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders, particularly of the immune system or endothelial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 146 as residues: Pro-41 to Cys-47, Phe-52 to Gly-59, Pro-62 to His-70. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in T-cells indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

10

30

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate igands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in 15 the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to 20 chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating 25 septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:33 and may have been publicly available prior to conception of

25

30

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1283 of SEQ ID NO:33, b is an integer of 15 to 1297, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:33, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

10 The translation product of this gene shares sequence homology with the human, mouse, and bovine dopamine hydroxylase which is thought to be important in the modification of dopamine, a neurotransmitter (See Genbank Accession Nos. gil30474, gil162965, and/or gil2358082; all references available through these accessions are hereby incorporated by reference herein). Preferred polypeptides of the invention comprise the following amino acid sequence: RQRSWNPGT 15 NCYHPNMPDAFLTCETVIFAWAIGGEGFSYPPHVGLSLGTPLDPHYVLLEVH YDNPTYEEGLIDNSGLRLFYTMDIRKYDAGVIEAGLWVSLFHTIPPGMPEF QSEGHCTLECLEEALEAEKPSGIHVFAVLLHAHLAGRGIRLRHFRKGKEMKL LAYDDDFDFNFQEFQYLKEEQTILPGDNLITECRYNTKDRAEMTWGGLSTR SEMCLSYLLYYPRINLTRCASIPDIMEQLQFIGVKEIYRPVTTWPFIIKSPKQYK 20 NLSFMDAMNKFKWTKKEGLSFNKLVLSLPVNVRCSKTDNAEWSIPRNDSIT SRYRKTL (SEQ ID NO: 303). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MCCWPLLLWGLLPGTAAGGSGRTYPHRTLLDSEGK YWLGWSQRGSQIAFRLQVRTAGYVGFGFSPTGAMASADIVVGGVAHGR PYLQDY FTNANRELKKDAQQDYHLEYAMENSTHTIIEFTRELHTCDINDKS ITDSTVRVIWAYHHE DAGEAGPKYHDSNRGTKSLRLLNPEKTSVLSTALPYF DLVNQDVPIPNKDTTYWCQMFKIPVFQEKHHVIKVEPVIQRGHESLVHHILL YQCSNNFNDSVPGIRARIAITPTCPMHSSPV KL (SEQ ID NO: 304). Polynucleotides encoding these polypeptides are also provided.

10

15

20

25

30

This gene is expressed primarily in brain, the pulmonary system, and to a lesser extent in kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological and behavioral disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., sputum, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 147 as residues: Ser-33 to Trp-38, Gly-40 to Gly-45, Asn-93 to Asp-105, Thr-128 to Thr-137, Glu-150 to Lys-167, Pro-197 to Tyr-203, Cys-242 to Asn-247, Ser-253 to Tyr-258, His-307 to Glu-314, Glu-357 to Gly-362, Trp-373 to Gln-378, Ser-402 to Glu-408. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to a protein involved in the modification of dopamine indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive

15

20

25

compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Alternatively, the homology to dopamine hydroxylase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-,hypoparathyroidism), hypothallamus, and testes. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:34 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2170 of SEQ ID NO:34, b is an integer of 15 to 2184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:34, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway.

10

15

20

25

30

The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates T-cells through the JakStat signal transduction pathway.

This gene is expressed in a variety of human normal and diseased tissues including breast, infant adrenal gland, skin tumor, colon, pituitary, Wilm's tumor, and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer and other proliferative disorders, afflicting endocrine or endothelial tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system or of breast and/or breast lymph nodes, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast, infant adrenal gland, skin turnor, colon, pituitary, and Wilm's turnor, and biological activity in activating the GAS promoter element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-,hypoparathyroidism), hypothallamus, and testes. Alternatively, the tissue distribution and biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the

diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells, (i.e., breast, skin and Wilm's tumors) indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:35 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 935 of SEQ ID NO:35, b is an integer of 15 to 949, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:35, and where b is greater than or equal to a + 14.

5

10

15

20

25

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: TGTFWSPRSQRRGCCGRRAPRPEAMENGAVYS PTTEEDPGPARGPRSGLAAYFFMGRLPLLRRVLKGLQLLLSLLAFICEEVVSQ CTLCGGLYFFEFVSCSAFLLSLLILIVYCTPFYERVDTTKVKSSDFYITLGTGCV FLLASIIFVSTHDRTSAEIAAIVFGFIASFMFLLDFITMLYEKRQESQLRKPENTT RAEALTEPLNA (SEQ ID NO: 305). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 3. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 3.

This gene is expressed primarily in dendritic cells, and to a lesser extent in melanocytes, fetal liver and spleen and several other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammation, and disorders of the hepatic and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic, hepatic, immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 149 as residues: Phe-63 to Ser-75, Thr-97 to Ser-102, Glu-128 to Arg-143. Polynucleotides encoding said polypeptides are also provided.

WO 00/06698

10

20

25

30

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in dendritic cells indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, 15 such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various would-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, raise

antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:36 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3324 of SEQ ID NO:36, b is an integer of 15 to 3338, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:36, and where b is greater than or equal to a + 14.

15

20

25

30

5

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ASAPRVMRGHLAGFPALSGLASVCLWATFSA QLPGPVAATSWTPAPLGCSAARSGPEKRLGTAAPGSAASLAQAGPGAPCRV LPVDPAPAALNVREPGWLGGLFDGALLQVLLNFLRKSTDVLMDTREAESLEV E (SEQ ID NO: 306).

In another embodiment polypeptides of the invention comprise the following amino acid sequence: NKLHSFPVFLSQLLLDRQLLHAPQTLPTPHCGGSSRPGP SHPPWLLIQLPCVHVALWQMLRDFSDSRITPSTLTTQPAAQTAAPAKDQES DIVGGEGILCDIAFLQEDHPLGVGGASAPSSRRELSRRGVHTQTLPEDGTLHG TPSSSFDCGIKYIISWPLAPGCDLPSLELSLVCKGVSSCMGFAAG (SEQ ID NO: 307). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in endothelial cells, lung, and fetal kidney, and to a lesser extent in epididymis, keratinocytes and cerebellum.

10

15

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cardiovascular diseases involving endothelial cell disturbances such as atheroschlerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 150 as residues: Arg-47 to Leu-54. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in endothelial cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating 20 disorders of endothelial cells such as atheroschlerosis, vasculitis, cardiovascular disease, and emphysema. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. The polypeptide may possess a wide range of undetected biological 25 activities. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction ofother cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmunediseases and allergy); regulation of haematopoiesis (e.g., for treatinganaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or

10

15

30

thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); antiinflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:37 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1549 of SEQ ID NO:37, b is an integer of 15 to 1563, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:37, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PGRPTRPTKNKVCVCLGMLFWAYPICVFIDSL SCQPCLWSTGATSHFNSPTTSPLFTLFMPCALAPNPFT QLGKLDDR (SEQ ID NO: 308). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, tumors or disorders of the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

10

a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 151 as residues: His-29 to Thr-34. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in meningima indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and 15 "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal 20 cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception, as well as disorders of the meninges such as meningioma and meningitis. In addition, the gene or gene 25 product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the 30 protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15

20

25

30

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:38 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1034 of SEQ ID NO:38, b is an integer of 15 to 1048, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:38, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

The translation product of this gene has been shown to encode a human brain specific mitochondrial carrier (Genbank Accession No. gil3851540lgblAAD04346.11 (AF078544); all references available through this accession are hereby incorporated herein by reference) which shares sequence homology with the human body weight disorder associated gene C5 product which is known to be differentially expressed in obese compared to lean mice (See GeneSeq Accession No. R91281). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with mitochondrial carriers proteins. Such activities are known in the art, some of which are described in Sanchis et al, J. Biol. Chem. 273:34611-34615 (1998), incorporated herein by reference.

Included in this invention as preferred domains are mitochondrial energy transfer protein (METP) domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Structurally, members of the family of mitochondrial energy transfer proteins consist of three tandem repeats of a domain of approximately one hundred residues. Each of these domains contains two transmembrane regions. As a signature pattern, we selected one of the most conserved regions in the repeated domain, located just after the first transmembrane region. To detect this widespread family of proteins, a consensus sequence was developed that contains the most conserved regions in the repeated domain. The consensus pattern is as follows: P.[DE].[LIVAT][RK].[LRH][LIVMFY][QMAIGV].

10

15

20

25

30

Preferred polypeptides of the invention comprise the following amino acid sequences: PVDLTKTRLQ (SEQ ID NO: 309) and PTDVLKIRMQ (SEQ ID NO: 310). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the METP domains of the sequence listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the METP domain. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the METP domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domain is characteristic of a signature specific to mitochondrial energy transfer proteins.

The gene encoding the disclosed cDNA is believed to reside on chromosome X. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome X.

This gene is expressed primarily in brain, and to a lesser extent, in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological and behavioral disorders and immune disorders and/or obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disgestive, immune, and nervous systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 152 as residues: Gln-189 to Gly-195. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to mitochondrial carrier proteins indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:39 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

30

formula of a-b, where a is any integer between 1 to 1416 of SEQ ID NO:39, b is an integer of 15 to 1430, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:39, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 30 5

Preferred polypeptides of the invention comprise the following amino acid sequence: MTFGSTISPTSTHASPSLGFCCSWLLEDLEEQLYCSAFEEAALTR RICNPTSCWLPLDMELLHRQVLALQTQRVLLGMWLRRAWDTWVSPRRVAP GSRCLLTASHPCTEKRRKASAXQRNLGYPLAMLCLLVLTGLSVLIVAIHILEL LIDEAAMPRGMQGTSLGQVSFSKLGSFGAVIQVVLIFYLMVSSVVGFYSSPLF RSLRPRWHDTAMTQIIGNCVCLLVLSSALPVFSRTLGLTRFDLLGDFGRFNWL GNFYIVFLYNAAFAGLTTLCLVKTFTAAVRAELIRAFGLDRLPLPVSGFPQAS RKTQHQ (SEQ ID NO: 311). Polynucleotides encoding such polypeptides are also provided.

15 This gene is expressed primarily in immune system tissues (e.g. resting Tcells, primary dendritic cells, and neutrophils, apoptotic T-cells) and umbilical vein. This gene is expressed to a lesser extent in the gastrointestinal tissue (e.g. small intestine, colon), brain (e.g. cerebellum, frontal cortex), aorta endothelial cells, skin tumor, embryonic tissue, thymus, and cancers (e.g. cheek, breast, synovial).

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for 25 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and gastrointestinal tract expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 153 as residues: Asp-21 to Ser-29. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells (e.g. T-cells, dendritic cells, neutrophils) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for 15 immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and 20 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in skin tumors and cancerous tissue (e.g. cheek, breast, synovial sarcoma) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in cellular sources such as embryonic tissue marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the

10

15

20

25

30

expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. The tissue distribution in cerebellum and frontal cortex indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

10

15

20

25

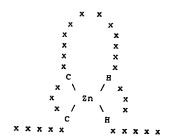
Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:40 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2089 of SEQ ID NO:40, b is an integer of 15 to 2103, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:40, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The polypeptide of this gene has been determined to have a zinc finger (Zinc finger, C2H2 type) domain at about amino acid position 16-50 of the amino acid sequence referenced in Table 1 for this gene. Therefore,

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: LCVCLVYLCMYGVCLCVIVCVSGVSLCLYVWGVSVC DCVSVFMCVCLCVIFCVYGKPRTEHYHSPHLAKQKAFREMCGRHDVSAAGIF QSYV (SEQ ID NO: 312). Polynucleotides encoding these polypeptides are also provided. 'Zinc finger' domains are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:

10



Many classes of zinc fingers are characterized according to the number and

positions of the histidine and cysteine residues involved in the zinc atom coordination.

In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc- dependent DNA or RNA binding property of some members of this class. Some of the proteins known to include C2H2-type zinc fingers are listed below. We have indicated, between brackets, the number of zinc finger regions found in each of these proteins; a '+' symbol indicates that only partial sequence data is available and that additional finger domains is present.

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 25 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, salivary gland related diseases, diseases of the mouth, and other digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above 30 tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., saliva, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 35 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 154 as residues: Gly-46 to His-54. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of salivary gland related diseases (mumps, calculi formation in ducts, sarcoidosis, facial palsy, tumors, Sjogrens Syndrome) and other digestive system disorders. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:41 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2335 of SEQ ID NO:41, b is an integer of 15 to 2349, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:41, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in fetal tissue (e.g. spleen, liver, brain), cancerous tissues (e.g. ovarian, colon, stomach, parathyroid) and to a lesser extent in immune cells and tissue (e.g. B-cells, T-cells, bone marrow), and reproductive organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly of the colon and ovaries, disorders of the developing fetus, neurodegenerative conditions, and immune system disorders.

10

15

20

25

30

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, reproductive, neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 155 as residues: Lys-35 to Lys-47. Polynucleotides encoding said polypeptides are also provided.

The expression of this gene within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in

15

- 20

25

30

modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (such as T-cells and B-cells) and immune tissues (bone marrow) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in parathyroid indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for

15

20

25

30

the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancrease (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-,hypoparathyroidism), hypothallamus, and 5 testes. Additionally, the tissue distribution in brain tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:42 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1545 of SEQ ID NO:42, b is an integer of 15 to 1559, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:42, and where b is greater than or equal to a + 14.

5

10

15

20

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in skin tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin disorders, particulary skin cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 156 as residues: Pro-38 to Gly-44, Phe-56 to Thr-64. Polynucleotides encoding said polypeptides are also provided.

10

15

20

25

The tissue distribution in skin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders including congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e.wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, uticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. Moreover, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, althletes foot, and ringworm). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:43 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1752 of SEQ ID NO:43, b is an integer of 15 to 1766, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:43, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

The translation product of this gene shares sequence homology with mitogeninduced prostate carcinoma (mouse) which is thought to be important in the etiology
of cancer. In this respect, this gene is mitogen-induced and/or involved in cell
proliferation.

10

15

20

25

Preferred polypeptides of the invention comprise the following amino acid sequence: GHMPYGWLTEIRAVYPAFDKNNPSNKLVSTSNTVTAAHIKKF
TFVCMALSLTLCFVMFWTPNVSEKILIDIIGVDFAFAELCVVPLRIFSFFPVPVT
VRAHLTGWLMTLKKTFVLAPSSVLRIIVLIASLVVLPYLGVHGATLGVGSLLA
GFVGESTMVAIAACYVYRKQKKKMENESATEGEDSAMTDMPPTEEVTDIVE
MREENE (SEQ ID NO: 313) and/or QVVFVAILLHSHLECREPLLIPILSLYMGA
LVRCTTLCLGYYKNIHDIIPDRSGPELGGDATIRKMLSFWWPLALILATQRISR
PIVNLFVSRDLGGSSAATEAVAILTATYPV (SEQ ID NO: 314). Polynucleotides
encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

This gene is expressed primarily in early infant and adult brain, retina, fetal tissue (e.g., liver, speen, whole embryo) and to a lesser extent in immune cells (e.g., monocytes and T-cells), colon, and parathyroid tumor tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, disorders of the immune system and nervous system.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic system (cancers), expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, neural,

cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic

epitopes shown in SEQ ID NO: 157 as residues: Arg-122 to Ser-139, Met-144 to Glu149. Polynucleotides encoding said polypeptides are also provided.

10

15

20

25

30

The tissue distribution and homology to mitogen induced prostate carcinoma (mouse) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study and treatment of cancers, including but not limited to the colon, parathyroid, and adrenal glands. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (T-cells, monocytes) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or

WO 00/06698

5

10

15

20

30

activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the

brain indicates it plays a role in normal neural function.

10

15

20

25

30

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:44 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2558 of SEQ ID NO:44, b is an integer of 15 to 2572, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:44, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in adult pulmonary tissue, umbilical vein, prostate, and fetal tissue (e.g.,heart).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

10

15

20

25

30

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 158 as residues: Arg-45 to Gly-51, Glu-75 to Asn-81. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in pulmonary tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of disorders associated with developing lungs, particularly in premature infants where the lungs are the last tissues to develop. Additionally, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and intervention of lung tumors, since the gene is involved in the regulation of cell division, particularly since it is expressed in fetal tissue. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in

10

15

25

30

proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:45 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 512 of SEQ ID NO:45, b is an integer of 15 to 526, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:45, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene is expressed primarily in adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fat metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

10

15

20

30

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 159 as residues: Pro-96 to Ser-106. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in adipose tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of obesity and other metabolic and endocrine conditions or disorders. Furthermore, the protein product of this gene may show utility in ameliorating conditions which occur secondary to aberrant fatty-acid metabolism (e.g. aberrant myelin sheath development), either directly or indirectly. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:46 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1018 of SEQ ID NO:46, b is an integer of 15 to 1032, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:46, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in adult brain tissue, testes, placenta, kidney, infant and fetal tissue (e.g.,liver, spleen, lung) and to a lesser extent in immune cells (e.g.,T-cells and neutrophils) and in cancerous tissues (e.g.,ovarian tumor, Hodgekins lymphoma, pancreas, T-cell).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

10

30

not limited to, CNS disorders, disorders of the testicles, cancer, particularly ovarian, pancreatic, T-cell, and Hodgekin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain,CNS, and testes expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, urogenital, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. 15 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, 20 trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product 25 in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders"

10

15

20

25

30

and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating. detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that 5 influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Additionally, the 10 tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to 15 identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play 20 related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. 25 Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:47 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2666 of SEQ ID NO:47, b is an integer of 15 to 2680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:47, and where b is greater than or equal to a + 14.

5

10

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

When tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates fibroblast cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation.

This gene is expressed primarily in endometrial stromal cells, endometrial tumors, keratinocytes, fetal tissue (e.g. liver, spleen) and to a lesser extent in endothelial cells and immune cells (e.g., T-cells).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial carcinoma and immune cells disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be

15

20

25

30

transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelyhood of implantation and development of a healthy embryo. In both cases this gene or its gene product could be administered at later stages of pregnancy to promote heathy development of the endometrium. Additionally, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endometrial carcinoma. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. The tissue distribution in immune cells such as helper Tcells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such

30

as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity 5 disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in 10 the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections 15 below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell 20 carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, uticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, althletes foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chrondomalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma,

dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, amd

15

20

25

30

chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:48 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1716 of SEQ ID NO:48, b is an integer of 15 to 1730, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:48, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in LNCAP cells (prostate cell line) and retina derived N2b5HR cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate cancer and eye disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urogenital and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative

10

15

20

25

30

to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 162 as residues: Asn-50 to Ser-57. Polynucleotides encoding said polypeptides are also provided.

The expression in prostate may indicate the gene or its products can be used in the disorders of the prostate, including inflammatory disorders, such as chronic prostatitis, granulomatous prostatitis and malacoplakia, prostatic hyperplasia and prostate neoplastic disorders, including adenocarcinoma, transitional cell carcinomas, ductal carcinomas, squamous cell carcinomas, or as hormones or factors with systemic or reproductive functions. The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma, retinochoroiditis, retinopathy and retinoschisis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:49 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1261 of SEQ ID NO:49, b is an integer of 15 to 1275, where both a and b correspond to the position's of nucleotide residues shown in SEQ ID NO:49, and where b is greater than or equal to a + 14.

15

20

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: RCCCRGCSCRARLCPPARSTAVAPECRGAHPSR AMRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTQRPCYKVIYFHD TSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID NO: 315). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGG TQRPCYKVIYFHDTSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID NO: 316). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle and human thyroid and to a lesser extent in amniotic cells and human endometrial stromal cells-treated with progesterone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, thyroid disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 163 as residues: Ser-75 to Leu-81. Polynucleotides encoding said polypeptides are also provided.

10

20

25

30

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endocrine disorders of the thyroid.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:50 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1748 of SEQ ID NO:50, b is an integer of 15 to 1762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:50, and where b is greater than or equal to a + 14.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene is expressed primarily in human testes tumor and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the testicles including but not limited to testicular cancerand immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and immune system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 164 as residues: His-31 to Gly-41. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in testes, particularly testicular tumors, indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. The tissue distribution in bone marrow indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

10

15

20

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:51 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2045 of SEQ ID NO:51, b is an integer of 15 to 2059, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:51, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

The translation product of this gene shares sequence homology with protocadherins, which are related to cadherin, and possess cell adhesive ability.

Cadherins are glycosylated integral membrane proteins that are involved in cell-cell adhesion.

This gene is expressed primarily in brain (infant, adult frontal lobe, manic depression tissue) and to a lesser extent in epididymus, healing groin wounds, ovary, adipocytes, and fetal tissue (e.g., kidney and retina).

WO 00/06698 PCT/US99/17130 112

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired male and female fertility, developmental disorders, fibrosis, and manic depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

30

15 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 165 as residues: Val-35 to Lys-41, Ser-68 to Gln-73, Glu-88 to Glu-93, Arg-156 to Gly-163, Ala-199 to Gly-206, Asp-216 to Ser-226, Thr-249 to Asn-254, Asp-339 to Pro-345, Ile-370 to Gly-379, Pro-429 to Glu-434, Arg-461 to Pro-466, Ala-475 to Thr-482, Pro-585 to Gly-593, Glu-631 to Gln-639, Pro-674 to Pro-682, Gln-715 to Gly-720, Ser-736 to Arg-742. Polynucleotides encoding -20 said polypeptides are also provided.

BLAST analysis reveals high homology to protocadherin sequences. These sequences are related to cadherin, and possess cell adhesive ability. Such proteins may have regulatory functions in the cell, as well as the cell-cell adhesive properties. Antibodies produced against these sequences are useful for modulating the binding 25 activity of these protocadherins, and can be used therapeutically. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease,

15

20

25

30

Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, 10 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Moreover, the expression within fetal tissue (e.g., kidney and retina) and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including blindness, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent WO 00/06698 PCT/US99/17130

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may 5 also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:52 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3268 of SEQ ID NO:52, b is an integer of 15 to 3282, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:52, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

10

15

20

25

Preferred polypeptides of the invention comprise the following amino acid sequence:

25

30

IRHEQQGEEDDEHARPLAESLLLAIADLLFCPDFTVQSHRRSTVDSAEDVHSL DSCEYIWEAGVGFAHSPQPNYIHDMNRMELLKLLLTCFSEAMYLPPAPESGS TNPWVQFFCSTENRHALPLFTSLLNTVCAYDPVGYGIPYNHLLFSDYREPLVE EAAQVLIVTLDHDSASSASPTVDGTTTGTAMDDADPPGPENLFVNYLSRIHRE EDFQFILKGIARLLSNPLLQTYLPNSTKKDPVPPGAASSLLEALRLQQEIPLLRA EEQRRPRHPCPHPLLPQRCPGRSV (SEQ ID NO: 317). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in brain, breast, breast cancer tissue and to a lesser extent in epididymus, amniotic cells, and embryo tissue.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired CNS function, male sterility, and breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential 15 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, male reproductive, cancerous and wounded tissues) or 20 bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 166 as residues: Pro-22 to Pro-31, Ser-38 to His-43, Asp-74 to Leu-79, Asp-113 to Glu-121, Leu-157 to Val-166, Ala-189 to Arg-196, Gln-206 to Arg-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particulary in the cerebellum, indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11,

15

20

25

30

15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. The expression in the breast tissue may indicate its uses in the diagnosis and/or treatment of breast neoplasia and breast cancers, such as fibroadenoma, pipillary carcinoma, ductal carcinoma, Paget's Disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma, as well as juvenile hypertrophy and gynecomastia, mastitis and abscess, duct ectasia, fat necrosis and fibrocystic diseases. Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may

10

15

20

25

30

show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating. detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:53 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

formula of a-b, where a is any integer between 1 to 1846 of SEQ ID NO:53, b is an integer of 15 to 1860, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:53, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 44

Contact of cells with supernatant expressing the product of this gene increases the permeability of monocytes to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells.

This gene is expressed primarily in CD34 positive cells derived from human cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; defects in hematopoietic stem and progenitor cells; susceptibility to chemotherapy and irradiation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 167 as residues: Ala-38 to Leu-59, Ala-63 to Thr-71, Lys-82 to Leu-91, Glu-97 to Ser-107, Gln-143 to Ala-149, Ile-153 to Leu-158, Ser-169 to Arg-182. Polynucleotides encoding said polypeptides are also provided.

10

15

20

25

30

Elevated expression of this gene product in CD34 positive hematopoietic cells indicates that it is expressed by early stem and progenitor cells of the hematopoietic lineages. Therefore, this may represent a soluble factor that is able to control the survival, proliferation, differentiation, or activation of all hematopoietic lineages, including stem and progenitor cells. Thus, it could be quite useful, for example, in ex vivo expansion of stem cell numbers for hematopoietic disorders or for cancer patients. Alternately, it may represent a factor that influences the hematopoietic microenvironment by affecting stromal cells that release other factors required for hematopoietic development. Additionally, the tissue distribution in CD34 positive cells also indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:54 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:54, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:54, and where b is greater than or equal to a + 14.

10

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

This gene is expressed primarily in breast and 12-week old human embryos and to a lesser extent in stomach cancer and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer; stomach cancer; embryonic defects: hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and/or treatment of a variety of disorders. Elevated expression of this gene product in stomach cancer indicates it is useful as a marker or therapeutic target for stomach cancer. Alternately, expression in breast tissue is influenced by the presence or absence of breast cancer tissue, and may thus also serve as a diagnostic marker for this cancer as well. Expression in the developing embryo may correlate with the normal development of human embryhos, and expression in the liver is involved in the regulation of normal liver function and/or liver regeneration.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:55 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

30

formula of a-b, where a is any integer between 1 to 1079 of SEQ ID NO:55, b is an integer of 15 to 1093, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:55, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in human hypothalamus derived from a patient with schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, schizophrenia; neurological disorders; impaired nervous system function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 169 as residues: Glu-34 to Trp-39. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particularly in the hypothalamus, indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal

15

20

30

cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:56 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 618 of SEQ ID NO:56, b is an integer of 15 to 632, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:56, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares sequence homology with human lecithin-cholesterol acyltransferase (LCAT), which catalyses the transfer of fatty acid from the sn-2 position of lecithin to the free hydroxyl group of cholesterol. Preferred polypeptides of the invention comprise the following amino acid sequence: RLVYN KTSRATQFPDGVDVRVPGFGKTFSLEFLDPSKSSVGSYFHTMVESLVGWGYT RGEDVRGAPYDWRRAPNENGPYFLALREMIEEMYQLYGGPVVLVAHSMGN MYTLYFLQRQPQAWKDKYIRAFVSLGAPWGGVAKTLRVLASGDNNRIPVIG

10

15

20

25

30

PLKIREQQRSAVSTSWLLPYNYTWSPEKVFVQTPTINYTLRDYRKFFQDIGFE DGWLMRQDTEGLVEATMPPGVQLHCLYGTGVPTPDSFYYESFPDRDPKICFG DGDGTVNLKSALQCQAWQSRQEHQVLLQELPGSEHIEMLANATTLAYLKRV LLGP (SEQ ID NO: 318). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in osteoblasts & dendritic cells and to a lesser extent in muscle and other hematopoietic cell lineages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; osteoporosis; osteopetrosis; muscle degeneration. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 170 as residues: Cys-65 to Ser-71. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution and homology to lecithin-cholesterol acyltransferase (LCAT) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. For example, artheroscelerosis is a pathological condition of mammals characterised by the accumulation of cholesterol in the arteries, which leads to heart disease, strokes, heart attacks and peripheral vascular disease. The enzyme could be used in a novel method of treating atherosclerosis, which involves increasing the level of LCAT activity, which then causes a decrease in the accumulation of cholesterol. The method and the

15

products can be used for the prophylaxis and treatment of atherosclerosis, and associated heart disease, myocardial infarction, stroke and peripheral vascular disease, as well as individuals suffering from Fish Eye Syndrome (caused by LCAT deficiency) or Classic LCAT Deficiency Syndrome. Alternately, elevated expression of this gene product in osteoblasts and hematopoietic cell lineages indicates that it may play additional roles in bone turnover, regulation of immune system function, and muscular function.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:57 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2673 of SEQ ID NO:57, b is an integer of 15 to 2687, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:57, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

When tested against HELA epithelial cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates epithelial cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in adult brain, infant brain, fibroblasts,

embryonic and fetal tissue (e.g., spleen, liver), placenta and to a lesser extent in
endocrine organs, cancerous colon and breast.

10

15

20

25

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, dementia, epilepsy, schizophrenia, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, endocrine system, and during development, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. In addition, the expression of this gene product in synovium (synovial sarcoma) would suggest a role in the detection and treatment of disorders and

conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chrondomalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and 5 chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). The tissue distribution in endocrine tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are 10 described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancrease (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, 15 hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-,hypoparathyroidism), hypothallamus, and testes. Additionally, the expression within fetal tissue, cancerous colon and breast, and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular 20 division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern 25 formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

30

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:58 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 605 of SEQ ID NO:58, b is an integer of 15 to 619, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:58, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

Preferred polypeptides of the invention comprise the following amino acid sequence or a subfragment thereof: MNKEDKVWNDCKGVNKLTNLEEQYIILIFQ NGLDPPANMVFESIINEIGIKNNISNFFAKIPFEEANGRLVACTRTYEESIKGSC GQKENKIKTVSFESKIQLRSKQEFQFFDEEEETGENHTIFIGPVEKLIVYPPPPA KGGISVTNEDLHCLNEGEFLNDVIIDFYLKYLVLEKLKKEDADRIHIFSSFFYK RLNQRERRNHETTNLSIQQKRHGRVKTWTRHVDIFEKDFIFVPLNEAAHWFL AVVCFPGLEKPKYEPNPHYHENAVIQKCSTVEDSCISSSASEMESCSQNSSAK

10

15

20

PVIKKMLNKKHCIAVIDSNPGQEESDPRYKRNICSVKYSVKKINHTASENEEF NKGESTSQKS (SEQ ID NO: 319). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in fetal tissue, stomach, brain, endometrial cells, and bone and to a lesser extent in prostate, retina, adipocytes, smooth muscle, and tumors of the endometrium, ovaries, and parathyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the endocrine system, ulcers, stomach cancer, epilepsy, schizophrenia, dementia, bone growth, developmental disorders and resorption. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system and neural systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.

25 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS,

10

15

20

25

30

psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Expression of this gene product in stomach tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of stomach cancer, and cancer in general. The expression within embryonic, fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in parathyroid tumor indicates polynucleotides and polypeptides corresponding to this

25

30

gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for 5 the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancrease (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-,hypoparathyroidism), hypothallamus, and testes. The tissue distribution in testes indicates that polynucleotides and 10 polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male 15 contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:59 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1364 of SEQ ID NO:59, b is an integer of 15 to 1378, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:59, and where b is greater than or equal to a + 14.

10

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

The translation product of this gene shares good protein homology with Xenopus NaDC-2 gene and a rabbit renal sodium/dicarboxylate cotransporter. The translation product of this gene also shares good homology with a rat placental protein which is a sodium-coupled high affinity dicarboxylate transporter. Therefore, it is likely that that the translated product encoded by this gene shares similar biological activity.

This gene is expressed primarily in the placenta and colon adenocarcinoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental abnormalities as well as failure to thrive anomalies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system and colon, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 173 as residues: Lys-166 to Gly-181. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human placenta and the shared homology of this translation product to a rat placental protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product is produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or

. 5

10

15

20

30

survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product is produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. The tissue distribution in colon tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders involving the colon. Expression of this gene product in colon tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of colon cancer, and cancer in general.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:60 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1112 of SEQ ID NO:60, b is an integer of 15 to 1126, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:60, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in the spinal cord.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, paralysis, neurologic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of

25

30

the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in spinal cord indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or 10 prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection. treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, 15 Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including 20 disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:61 and may have been publicly available prior to conception of

10

15

20

25

30

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2064 of SEQ ID NO:61, b is an integer of 15 to 2078, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:61, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene is expressed primarily in keratinocytes, brain, fetal tissues, pericardium, stomach, and cancerous tissues (e.g., stomach, adrenals, parathyroid, germ cell, colon, breast).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin disorders, neurodegenerative and developmental disorders, heart disease, and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and gastrointestinal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of

Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, 10 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and 15 "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis 20 fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e.wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, uticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, althletes foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chrondomalacia and 30 inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, amd chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita,

15

20

25

30

familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). The expression within fetal tissue (e.g., spleen and liver) and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Additionally, the tissue distribution in the pericardium of the heart indicates that the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the

10

20

25

30

protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:62 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 748 of SEQ ID NO:62, b is an integer of 15 to 762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:62, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene is expressed primarily in the brain and in cartilage and to a lesser extent in the retina, activated T-cells, pineal gland, the lungs, and in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological diseases, such as epilepsy and dementia, osteoarthritis, retinopathies, hematopoietic diseases, emphysema, and lung cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurologic system, cartilage and musculature, vision, the hematopoietic system, and the pulmonary system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 176 as residues: Arg-34 to Cys-44. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in T-cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such

15

20

25

30

as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The expression of this gene product in synovium would suggest a role in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chrondomalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). Additionally, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have

15

20

25

applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:63 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1080 of SEQ ID NO:63, b is an integer of 15 to 1094, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:63, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

This gene is expressed primarily in umbilical vein endothelial cells induced by IL-4.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

15

20

25

30

not limited to, angiogenesis, inflammatory disorders, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the angiogenic and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in endothelial cells indicates polynucleotides and polypeptides corresponding to this gene are useful in the detection, treatment, and/or prevention of vascular conditions, which include, but are not limited to, microvascular disease, vascular leak syndrome, aneurysm, stroke, atherosclerosis, arteriosclerosis, or embolism. For example, this gene product may represent a soluble factor produced by smooth muscle that regulates the innervation of organs or regulates the survival of neighboring neurons. Likewise, it is involved in controlling the digestive process, and such actions as peristalsis. Similarly, it is involved in controlling the vasculature in areas where smooth muscle surrounds the endothelium of blood vessels. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human

10

15

20

immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:64 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1347 of SEQ ID NO:64, b is an integer of 15 to 1361, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:64, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene is expressed primarily in both normal and cancerous pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, diabetes, gastrointestinal disorders, and pancreatic cancer. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
a number of disorders of the above tissues or cells, particularly of the digestive and
blood systems, expression of this gene at significantly higher or lower levels is
routinely detected in certain tissues or cell types (e.g., cancerous and wounded

10

15

20

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in pancreas indicates that the protein products of this gene are useful as a therapeutic and/or diagnostic agent for pancreatic disorders and disorders of the endocrine and exocrine system, including but not limited to diabetes, blood disorders, pancreatic cancer, gastrointestinal diseases, hormomal imbalance, autoimmune disorders, cystic fibrosis, pancreatitis, and gallstones. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:65 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 933 of SEQ ID NO:65, b is an integer of 15 to 947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:65, and where b is greater than or equal to a + 14.

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

The translation product of this gene shares sequence homology with oxidoreductase. Preferred polypeptides of the invention comprise the following amino acid sequence: MSPLSAARAALRVYAVGAAVILAQLLRRCRGGFLEPVXPPRP DRVAIVTGGTDGIGYSTANIWRDLGMHVIIAGNNDSKAKQVVSKIKEETLND KVEFLYCDLASMTSIRQFVQKFKMKKIPLHVLINNAGVMMVPQRKTRDGFEE HFGLNYLGHFLLTNLLLDTLKESGSPGHSARVVTVSSATHYVAELNMDDLOS

20

25

30

SACYSPHAAYAQSKLALVLFTYHLQRLLAAEGSHVTANVVDPGVVNTDXYK HVFWATRLAKKLLGWLLFKTPDEGAWTSIYAAVTPELEGVGGRYLYNEKET KSLHVTYNQKLQQQLWSKSCEMTGVLDVTL (SEQ ID NO: 320). The mature form of this protein begins at residue 32. Thus, polypeptides comprising residues 2-330 and 32-330 of the sequence shown above are also provided. Polynucleotides encoding such polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MSPLSAARAALRVYAVGAAVILAQLLRRCRGGFLEP VXPPRPDRVAIVTGGTDGIG YSTANIWRDLACMLS (SEQ ID NO: 321).

10 Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in breast cancer cells, osteoclastoma, wilm's tumor, thymus stromal cells, and T cell helper I.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, e.g., breast cancer, osteoclastoma, and wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, kidney, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, breast milk, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast cancer tissue, combined with the homology to oxidoreductase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer, particularly, breast cancer, osteoclastoma, and wilm's tumor. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative

10

15

20

25

30

conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating. detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:66 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1362 of SEQ ID NO:66, b is an

10

15

20

integer of 15 to 1376, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:66, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in monocytes, T cell helper II and B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly B-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 180 as residues: Asp-30 to Val-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in monocytes, T cell helper, and B cell lymphoma cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell lymphoma. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

5

10

15

20

25

30

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities. such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:67 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2420 of SEQ ID NO:67, b is an integer of 15 to 2434, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:67, and where b is greater than or equal to a + 14.

10

15

20

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene is expressed primarily in human lung cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, pulmonary diseases and/or disorders, particularly cancers of the lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary lavage, pulmonary surfactant, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 181 as residues: Phe-39 to Asp-45. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in lung cancer tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as ARDS, cystic fibrosis, and cancer, particularly lung cancer. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell

death, as occurs in the development of some cancers, or in failure to control the extent
of cell death, as is believed to occur in acquired immunodeficiency and certain
neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

25

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, 5 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in 10 proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as. 15 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:68 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1072 of SEQ ID NO:68, b is an integer of 15 to 1086, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:68, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in larynx carcinoma and early stage human 30 lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

15

20

25

30

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, gastrointestinal, and pulmonary diseases and/or disorders, particularly larynx carcinoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, gastrointestinal, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, pulmonary lavage, sputum, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 182 as residues: His-42 to Lys-49. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in larynx carcinoma and early stage human lung indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating immune system disorders such as cancer, particularly larynx carcinoma. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

15

20

30

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:69 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1248 of SEQ ID NO:69, b is an integer of 15 to 1262, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:69, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 60

Preferred polypeptides of the invention comprise the following amino acid sequence: MEVTTEDTSRTDVSEPATSGGAADGVTSIAPTAVASSTTAASITTA ASSMTVASSAPTTAASSTTVASIAPTTTASSMTAASSTPMTLALPAPTSTXTGR TPSTTATGHPSLSTALAQVPKSSALPRTATLATLATRAQTVATTANTSSPMST RPSPSKHMPSDTAASPVPPMXPQAQGPISQVSVDQPVVNTTXKSTXMPSNTT XEPLTQAVVDKTLLLVVLLLGVTLFITVLVLFALQAYESYKKKDYTQVDYLI

15

20

25

NGMYADSEM (SEQ ID NO: 322). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARCPELPGLRCRPRPRAGPQAPSYCPRATRPPG ACCARMRLLLEWRVYLRLTCATKDGMARECPTTWLSPPAKPDFAQRHSVK PTALQGGRWSRLGASP (SEQ ID NO: 323). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in adipocytes, osteoblasts, cerebellum, hypothalamus and Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic, skeletal, neural, and immune diseases and/or disorders, particularly Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, skeletal, neural, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 183 as residues: Pro-33 to Gln-40, Gly-51 to Arg-56. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in Hodgkin's lymphoma cells indicates that

30 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as cancer, particularly Hodgkin's lymphoma. The secreted protein can also be used to determine biological activity, to

raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:70 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1628 of SEQ ID NO:70, b is an integer of 15 to 1642, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:70, and where b is greater than or equal to a + 14.

25

10

15

20

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

The translation product of this gene shares sequence homology with polypeptide in the cystatin family. Cystatin polypeptides are cysteine protease inhibitors. For an analysis of the composition of several members of the cystatin family see Gene (1987) 61(3):329-338, incorporated herein by reference. The cystatin activity of polypeptides encoded by this gene is measured by several assays known in the art including assays described in coowned, copending US Patent Application Serial No. 08/744,138, incorporated herein by reference. Preferred polypeptides of the invention comprise the following amino acid sequence: LPATVEFAVHTFNQQSKD YYAYRLGHILNSWKEQVESKTVFSMELLLGRTRCGKFEDDIDNCHFQESTEL NNTFTCFFTISTRPWMTQFSLLNKTC (SEQ ID NO: 324). Fragments of such polypeptides having cystatin activity (cysteine protease inhibitory activity are particularly preferred). Polynucleotides encoding such polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LLWARGLGRAKSAVPTVST MLGLPWKGGLS WALLLLLGSQILLIYAWHFHEQRDCDEHNVMARYLPATVEFAVHTFNQQS KDYYAYRLGHILNSWKEQVESKTVFSMELLLGRTRCGKFEDDIDNCHFQE STELNNTFTCFFTISTRPWMTQFSLLNK TCLEGFH (SEQ ID NO: 325). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in testes and epididiymus. For a review of a cystatin showing testes- specific expression see Mol. Endocrinol. (1992 Oct.) 6(10):1653-1664, incorporated herein by reference.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can been detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes

15

20

25

30

in rheumatoid arthritis, purulent bronchiectasis and periodontitis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, seminal fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 184 as residues: Phe-31 to Asp-38, Asn-59 to Tyr-65, Ser-76 to Glu-82, Thr-96 to Cys-108, Gln-111 to Asn-118. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in testes and epididiymus, combined with the homology to cystatins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

15

20

30

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Cysteine proteinase inhibitors of the cystatin superfamily are ubiquitous in the body and are generally tight-binding inhibitors of papain-like cysteine proteinases, such as cathepsins B, H, L, S, and K. They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can been detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes in rheumatoid arthritis, purulent bronchiectasis and periodontitis, which indicates that a tight cystatin regulation is a necessity in the normal state. A deficiency state in which the levels of the intracellular cystatin, cystatin B, are lowered due to mutations has recently been shown to segregate with a form of progressive myoclonus epilepsy, which points to additional specialized functions of cystatins. Moreover, results showing that chicken cystatin inhibits polio virus replication, human cystatin C inhibits corona- and herpes simplex virus

10

15

20

25

30

replication, and human cystatin A inhibits rhabdovirus-induced apoptosis in cell cultures indicates that cystatins play additional roles in the human defense system. The cystatins constitute a superfamily of evolutionarily related proteins, all composed of at least one 100-120 residue domain with conserved sequence motifs.

The previously well characterized single-domain human members of this superfamily could be grouped in two protein families. The Family 1 members, cystatins (or stefins) A and B, contain approximately 100 amino acid residues, lack disulfide bridges, and are not synthesized as preproteins with signal peptides. The Family 2 cystatins (cystatins C, D, S, SN, and SA) are secreted proteins of approx. 120 amino acid residues (Mr 13,000-14,000) and have two characteristic intrachain disulfide bonds. Recently, we identified an additional human cystatin superfamily member by EST1 sequencing in epithelial cell derived cDNA libraries which we named cystatin E. The same cystatin was independently discovered by differential display experiments as a mRNA species down-regulated in breast tumor tissue, but present in the surrounding epithelium and reported under the name cystatin M. Cystatin E/M is an atypical, secreted low-Mr cystatin in that it is a glycoprotein and just shows 30-35% sequence identity in alignments with the human Family 2 cystatins, which shows that additional cystatin families are yet to be identified. The cystatin E/M gene has been localized to chromosome 2, whereas all human Family 2 cystatin genes are clustered on the short arm of chromosome 20, which further stresses that cystatin E/M is just distantly related to the other secreted human low-Mr cystatins. It is believed therefore, that polypeptides encoded by this gene are useful in diagnosing and treating disease consistent with the aforementioned conditions in which cystatins are implicated.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:71 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 907 of SEQ ID NO:71, b is an

10

15

20

25

30

integer of 15 to 921, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:71, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

The translation product of this gene shares sequence homology with Neutrophil Gelatinase-Associated Lipocalin which is thought to be important in immune regulation (See Genbank and Geneseq Accession Nos. emblCAA58127.1, and US5627034, respectively; all references and information available through these accessions are hereby incorporated herein by reference; for example, Biochem. Biophys. Res. Commun. 202 (3), 1468-1475 (1994), and FEBS Lett. 314 (3), 386-388 (1992)).

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LEQKLELHRGGGRSRTSGSPGLQEFGTREERGE GEQRTGREFSGNGGRAVEAARMRLLCGLWLWLSLLKVLQAQTPTPLPLPP PMQSFQGNQFQGEWFVLGLAGNSFRPEHRALLNAFTATFELSDDGRFEVWN AMTRGQHCDTWSYVLIPAAQPGQFTVDHGVGRSWLLPPGTLDQFICLGRAQ GLSDDNIVFPDVTGXALDL XSLPWVAAPA (SEQ ID NO: 326). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in epididiymus and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive and skeletal diseases and/or disorders, particularly cancers such as osteoclastoma testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, skeletal, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, urine, synovial fluid and spinal fluid)

or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 185 as residues: Met-82 to Thr-90. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididiymus and homology to neutrophil gelatinaseassociated lipocalin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of skin diseases and immune system 10 disorders such as cancer, particularly osteoclastoma. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" 15 sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); 20 regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction 25 etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or 30 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

20

25

30

related to SEQ ID NO:72 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 892 of SEQ ID NO:72, b is an integer of 15 to 906, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:72, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 63

The translation product of this gene was shown to have homology to colipase which plays an essential role in the intestinal fat digestion by anchoring lipase on lipid/water interfaces in the presence of bile salts (See Genbank Accession No. gblAAA03513.1; all references and information available through this accession are hereby incorporated by reference herein).

This gene is expressed primarily in epididiymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive diseases and/or disorders, particularly epididiymus-related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, bile, chyme, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 186 as residues: Ile-40 to Cys-49, Arg-52 to Cys-57,

15

20

25

30

Ser-94 to Trp-99, Gly-105 to Gly-111. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididiymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system diseases and disorders of the epididiymus. Polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:73 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 666 of SEQ ID NO:73, b is an integer of 15 to 680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:73, and where b is greater than or equal to a + 14.

10

15

20

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 64

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: MCVCERKRGREKEGGVTPTMTSNFPFCTLILGI AQAQACPGCPGDWPGLGSGVGEGLHHIRTCRTPIPCSPPAPAAACLGSGH ARLPCVLRLWPVPANLSSPFRLEALHCSFWSSPLLPAPHLAFFGFRDLLTDFL LAACLLTFQKTPLELPMAVVHLLVATPCYQMLDNLPLPSAAAN WC (SEQ ID NO: 327). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in melanocytes and placenta and to a lesser extent in bone marrow and many cells of the immune system, including B-cells, dendritic cells, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin cancer and disorders of the reproductive and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., reproductive tissue, hematopoietic tissue, melanocytes and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in melanocytes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the skin, the reproductive system, and the immune system, particularly cancers. Representative uses are described in the "Biological Activity",

WO 00/06698

10

15

20

25

30

"Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, uticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae. purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas. impetigo, tinea, althletes foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chrondomalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, amd chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:74 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

25

30

formula of a-b, where a is any integer between 1 to 1619 of SEQ ID NO:74, b is an integer of 15 to 1633, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:74, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Preferred polypeptides of the invention comprise the following amino acid sequence: YLWGRPRLRMRAGTSPSAPWGEKREKLGHKLPVALQGYHPWIL LECTVFWARVVLACFSLYLIRGPNCINRQPEPTYQKACNLDCSSDFGQER APAWELLGPESEQRLREYTAQGLQSLASSHRWRQFKTEGKMRGGASPLPWLI CFW LCSYKGSDNSLKPVVPGPTLCPQSLVSPSVHPSTRSASLGRHRAEAA (SEQ ID NO: 328). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: MPGILAGIPVKDLCLSLLQGFRLLLLCVCPGWL SGWMGGQKGSPRIVDIG (SEQ ID NO: 329). Polynucleotides encoding these polypeptides are also provided. This gene maps to chromosome 15, accordingly, polynucleotides of the invention is used in linkage analysis as a marker for chromosome 15.

This gene is expressed primarily in brain and breast and to a lesser extent in the liver, pancreas, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting the brain and CNS, the reproductive system, or the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, the reproductive system, and the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., brain and other tissue of the nervous system, mammary tissue, eendocrine tissue, hepatic tissue, reproductive tissue, cells

10

15

and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 188 as residues: Met-37 to Ser-43. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the central nervous system, the reproductive system, and the immune system, including cancers. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,

neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or
survival. Furthermore, the protein may also be used to determine biological activity,
to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to
identify agents that modulate their interactions, in addition to its use as a nutritional
supplement. Protein, as well as, antibodies directed against the protein may show

utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

15

related to SEQ ID NO:75 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1008 of SEQ ID NO:75, b is an integer of 15 to 1022, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:75, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 66

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AKGEERKEAFSLKMVQLSSEPISFGLMYLYLGV FFHLIYPGALSITTLGKHSHPFFTAEQNSTVWMEHTLFHQSPVASHLVCFQSF AFSE (SEQ ID NO: 330). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the brain and the immune system, in particular T-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting the brain, such as Alzheimer's or disorders affecting the immune system, such as AIDS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential 25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and CNS and the immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, cells and 30 tissue of the immune system. and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

10

15

20

25

30

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the brain and CNS or disorders affecting the immune system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:76 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1170 of SEQ ID NO:76, b is an

10

15

20

25

30

integer of 15 to 1184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:76, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

The translation product of this gene shares sequence homology with penaeidin-2 which is thought to be a members of a new family of antimicrobial peptides from the hemolymph of shrimps Penaeus vannamei. The molecules display antimicrobial activity against fungi and bacteria with a predominant activity against Gram-positive bacteria.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPAHPASPPLMTLSLQLAELVHFVCAFQSQWT GVYPMMPPLKPTEPLCFA CVPCRV (SEQ ID NO: 331). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly disorders affecting the spleen, including bacterial and fungal infections. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoetic and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, and cells and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

The tissue distribution in spleen and homology to the penaeidin family of antibiotics indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the spleen, especially fungal and bacterial infections. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

15

20

related to SEQ ID NO:77 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 298 of SEQ ID NO:77, b is an integer of 15 to 312, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:77, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 68

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating immune and hematopoietic cells and tissue cell types. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

Moreover, when tested in TF-1 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: c-fos, c-jun, egr-1, b561, bcl-2, CD40, cyclin D2, GADPH, ICER, MAD3, p21, STAT3, ID3, and STAT-1. When tested in U937 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: egr2, MKP1, ATF3, B562, cyclin D, cyclin D2, GATA3, MAD3, p21, TGF, DHFR, and JAK3. Based upon these results, it is anticipated that polynucleotides and polypeptides corresponding to this gene are useful as agonists or antagonists of the above referenced genes. Such activity is useful in therapeutic and/or diagnostic applications as referenced and more specifically discussed elsewere herein.

10

15

20

25

30

In specific embodiments, polpeptides of the invention comprise the sequence:MLLEVYGDSISVTVAIPL (SEQ ID NO:332),
MHSPCQSKAADGLGKSETE (SEQ ID NO: 333), and/or MLKSLGLSTN (SEQ ID NO: 334). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AQRLAEECFYMLLEVYGDSISVTVAIPLMHSP CQSKAADGLGKSETEMLKSLGLSTNMSPFHLLGLKVFLTWALTLAQICLY FFEVQPLGLLALNFFCTATAGLKELCMHPPSLAFTPEFHTSLSPLAIPSFCGTS VSLSNSHTIPLSLYLPFPSKSRMPDTLHLLVHSLPLVHSQVLPVKDVTIEWPLC QRCLGSTCH Q (SEQ ID NO: 335). Polynucleotides encoding these polypeptides are also provided.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 11 - 27 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 28 to 143 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

This gene is expressed primarily in neutrophils and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders and/or diseases affecting the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

10

15

20

25

30

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 191 as residues: Pro-97 to Asp-104. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in neutrophils and T-cells, combined with the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their

10

20

25

30

interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:78 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1356 of SEQ ID NO:78, b is an integer of 15 to 1370, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:78, and where b is greater than or equal to a + 14.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 69

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEVQVSLFQMFCFSSIFCSH EHTHLPGTFWLFLFLILPPSCPCFLPFSLAIETVRWPCWHHPTSFELCY PGTSIYYASRGGPXPNSEX (SEQ ID NO: 336). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases and/or disorders affecting the immune system, and neutrophils in particular. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., blood cells,

10

15

20

25

30

and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system and neutrophils in particular. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

10

15

20

25

30

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:79 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 354 of SEQ ID NO:79, b is an integer of 15 to 368, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:79, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: XNXKSPLTIGNKSWSSTAVAAALELVDPPGCR NSARDSPELVHLGKGRPRKLMTYLFCSSISLLLLKVHSSGHQDIRKAKSKVP RLLIIQCPQQRE (SEQ ID NO: 337). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting smooth muscle tissue, particularly vascular conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of smooth muscle tissue expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

10

15

20

25

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 193 as residues: Ser-18 to Val-31. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution primarily in smooth muscle indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting smooth muscle tissue. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:80 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1074 of SEQ ID NO:80, b is an integer of 15 to 1088, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:80, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by

10

20

25

30

27.

the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPEENLSPSTPSQMPTIWVKLCLLQVCHGLFP LLKHWSQPMPLCVTLAPVSYWL (SEQ ID NO: 338). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in fetal heart, smooth muscle, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, muscular, vascular, or neural diseases and/or disorders, particularly defects or injury to cardiac muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscular, vascular, neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal heart indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating defects to the heart either due to injury or congenital defects. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection.

10

15

20

treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,
Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating
diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal
cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia,
mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder,
learning disabilities, ALS, psychoses, autism, and altered behaviors, including
disorders in feeding, sleep patterns, balance, and perception. In addition, elevated
expression of this gene product in regions of the brain indicates it plays a role in
normal neural function. Furthermore, the protein may also be used to determine
biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or
receptors, to identify agents that modulate their interactions, in addition to its use as a
nutritional supplement. Protein, as well as, antibodies directed against the protein may
show utility as a tumor marker and/or immunotherapy targets for the above listed
tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:81 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1848 of SEQ ID NO:81, b is an integer of 15 to 1862, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:81, and where b is greater than or equal to a + 14.

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translation product of this gene shares sequence homology with adipose complement related protein which is thought to be important in regulating energy metabolism, insulin levels and fat stores. Moreover, the protein product of this gene has also been shown to have homology to the complement subcomponent C1Q Achain precursor and HP-25 protein (See Genbank and Geneseq Accession Nos. emblCAA41664.1, dbjlBAA02352.1, and W98013; all references and information

10

15

20

25

30

available through this accession are hereby incorporated by reference herein). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with complement proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PRVRKEPEAMQWLRVRESPGEATGHRVTMG TAALGPVWAALLLFLLMCEIPMVELTFDRAVASDCQRCCDSEDPLDPAHVSS ASSSGRPHALPEIRPYINITILKGDKGDPGPMGLPGYMGREGPQGEPGPQGSK GDKGEMGSPGAPCQKRFFAFSVGRKTALHSGEDFQTLLFERVFVNLDGC FDMATGQFAAPLRGIYFFSLNVHSWNYKETYVHIMHNQKEAVILYAQPS ERSIMQSQSVMLDLAYGDRVWVRLFKRQRENAIYSNDFDTYITFSGHLIKA EDD (SEQ ID NO: 339). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta and, fetal kidney, and umbilical vein and to a lesser extent in fetal heart, fetal liver/spleen, microvascular endothelial cells and cancers of the lung and pharynx.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, vascular, renal, and reproductive diseases and/or disorders, particularly cancers of the lung and pharynx. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, renal, reproductive, immune, hematpoietic, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 195 as residues: Asp-36 to Asp-48, Ser-57 to His-62, Lys-77 to Gly-84, Met-92 to Gly-114, Gln-203 to Ile-209, Lys-231 to Tyr-239. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in pharynx or lung, combined with the homology to adipose complement related proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancers of the pharynx or lung by modifying the metabolic balance in such tissues. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:82 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1604 of SEQ ID NO:82, b is an integer of 15 to 1618, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:82, and where b is greater than or equal to a + 14.

10

15

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The translation product of this gene shares sequence homology with a hypothetical 54.7 kD protein (F37A4.1) from Caenorhabditis elegans (SwissProt locus YPT1_CAEEL, accession P41879). The protein product of this gene also has homology to the human NG26 which is thought to contain a human major histocompatibility complex class III and is involved in T-cell maturation (See Genbank Accession No. gblAAD18079.11 (AF129756); all references and information available through this accession are hereby incorporated by reference herein; for example, J. Neurochem. 69 (6), 2516-2528 (1997)). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with nitric oxide synthase proteins.

Preferred polypeptides of the invention comprise the following amino acid sequence: MLYPGSVYLLQKALMPVLLQGQARLVEECNGRRAKLLACDGNE IDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTPLEAGYSVLGWNHP GFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSIGGFTATWAA MSYPDVSAMILDASFDDLVPLALKVMPDSWRGLVTRTVRQHLNLNNAEQLC RYQGPVLLIRRTKDEIITTTVPEDIMSNRGNDLLLKLLQHRYPRVMAEEGLRV VRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGPDFPWSVGEDMS ADGRRQLALFLARKHLHNFEATHCTPLPAQNFQMPWHL (SEQ ID NO: 340).

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VCPKWCRFLTMLGHCCYFWQVWPASEALAA GPTPSTGSSSPSWKQHIGTSLQKTRGSLPTTTLTSGAGQSTSTGKNPAAGR SLEGALPAGVWPCFAQSPCTGGQQTP SSTGLRSCLVRSPATWWRTP (SEQ ID NO: 341). Polynucleotides encoding these polypeptides are also provided.

Polynucleotides encoding such polypeptides are also provided.

Preferred polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEIYREXDSERAPASVPETPTAVTAPHSSSWDTYYQ

30 PRALEKHADSILALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAG
TLLLLAGVACXRGIGRWTNPQYRQFITILEATHRNQSSENKRQLANYNFD
FRSWPVDFHWEEPSSRKESRGGPSRRGVALLRPEPLHRGTADTLLNRVKKL

20

25

30

PCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQGQARLVEECNGRRAK LLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTPLEAGYS VLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSI GGFTATWAAMSYPDVSAMILDASFDDLVPLALKVMPDSWRGLVTRTVRQ HLNLNNAEQLCRYQGPVLLIRRTKDEIITTTVPEDIMSNRGNDLLLKLLQHRY PRVMAEEGLRVVRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGP DFPWSVGEDMSADGRRQLALFLARKHLHNFEATHCT PLPAQNFQMPWHL (SEQ ID NO: 342). Polynucleotides encoding these polypeptides are also provided. A preferred polypeptide variant of the invention comprises the following amino acid sequence: HERAXGPSRGHGELLSCVLGPRLYKIYRERDSERAPASVPETPTA10 VTAPHSSSWDTYYQP RALEKHADSILALASVFWSISYYSSPFAFFYLYRKGY LSLSKVVPFSHYAGTLLLLAGV ACSEALAAGPTPSTGSSSPSWKQHIGTSLQ KTRGSLPTTTLTSGAGQSTSTGKNPAAGRSLEGALPAGVWPCFAQSPCTGGQQTPSSTGL RSCLVRSPATWWRTP (SEQ ID NO: 343). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in cerebellum, pituitary, fetal liver, and primary dendritic cells and to a lesser extent in in a wide range of tissues and developmental stages (i.e. fetal and adult tissue, etc.).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural, developmental, and immune diseases and/or disorders, particularly those involving self recognition and T- anc B-cell maturation, and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural or hormonal system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, developmental, immune, hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., serum,

WO 00/06698

5

10

15

20

25

30

plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 196 as residues: Thr-23 to Lys-34, Leu-41 to Ser-47, Ala-57 to Ala-68, Pro-89 to Gly-101, Pro-110 to Pro-117. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in developmental and immune cells, combined with the homology to the human major histocompatibility complex class III region, indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and other proliferative disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and

15

25

30

in the differentiation and/or proliferation of various cell types. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:83 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2020 of SEQ ID NO:83, b is an integer of 15 to 2034, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:83, and where b is greater than or equal to a + 14.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 74

The translation product of this gene shares sequence homology with the br-1 protein from the snail nervous system (EMBL HPBR1GENE) which codes for nitric oxide synthetase and which is thought to be important in mediating a variety of cellular responses, including vasodilation. Preferred polypeptides of the invention comprise the following amino acid sequence: MFKRHQRLKKDSTQAEEDLSEQ EQNQLNVLKKHGYVVGRVGRTFLYSEEQKDNIPFEFDADSLAFDMENDPVM GTHKSTKQVELTAQDVKDAHWFYDTPGITKENCILNLLTEKEVNIVLPTQSIV PRTFVLKPGMVLFLGAIGRIDFLQGNQSAWFTVVASNILPVHITSLDRADALY QKHAGHTLLQIPMGGKERMAGFPPLVAEDIMLKEGLGASEAVADIKFSSAG WVSVTPNFKDRLHLRGYTPEGTVLTVRPPLLPYIVNIKGQRIKKSVAYKTKKP PSLMYNVRKKKGKINV (SEQ ID NO: 344). Polynucleotides encoding such polypeptides are also provided.

15

20

25

30

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MLPARLPFRLLSLFLRGSAPTAARHGLREPLLERRCAA ASSFQHSSSLGRELPYDPVDTEGFGEGGDMQERFLFPEYILDPEPQPTREKQL QELQQQEEEERQRQQRREERRQQNLRARSREHPVVGHPDPALPPSGVNCS GCGAXLHCQDAGVPGYLPREKFLRTAEADGGLARTVCQRCWLLSHHRRALR LQVSREQYLELVSAALRXPGPSLVLYMVDLLDLPDALLPDLPALVGPKQLIV LGNKVDLLPQDAPGYRQRLRERLWEDCARAGLLLAPGTKGHSAPSRTSHR TGRIRIRRTGPAQWSGTCG (SEQ ID NO: 345). Polynucleotides encoding these polypeptides are also provided.

When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid cells through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in early stage human brain, smooth muscle, and endometrial tumor and to a lesser extent in a variety of tissues representing many organs and developmental states.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cardiovascular, vascular, and neural diseases and/or disorders, particularly congestive heart disease and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and neural systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, vascular, neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

10

urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 197 as residues: Phe-42 to Leu-48, Pro-53 to Asp-58, Pro-81 to Glu-123, Asp-256 to Trp-269, Gly-282 to Ser-306, Arg-333 to Gly-339, Arg-403 to Gln-425, Ser-446 to Asn-452, His-475 to Gln-480, Gly-592 to Met-597, Pro-635 to His-642, Lys-667 to Lys-672, Lys-678 to Ser-684. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in smooth muscle and vascular tissues, combined with the homology to nitric oxide synthetase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of congestive heart failure and neurological degenerative disorders. polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or 15 prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, 20 treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including 25 disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,

neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or
survival. Moreover, the protein is useful in the detection, treatment, and/or prevention
of a variety of vascular disorders and conditions, which include, but are not limited to

10

15

20

25

30

: . _ _

miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis.

Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement.

Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:84 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2226 of SEQ ID NO:84, b is an integer of 15 to 2240, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:84, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with the human KE04p, in addition to an unidentifed C.elegans gene.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 9 - 25 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 to 8 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PSFRRERVETGGGGPVTHGTEGPFLPLPGGTRM NMTQARVLVAAVVGLVAVLLYASIHKIEEGHLAVYYRGGALLTSPSGPGYH

20

25

IMLPFITTFRSVQTTLQTDEVKNVPCGTSGGVMIYIDRIEVVNMLAPYAVFDIV RNYTADYDKTLIFNKIHHELNQFCSAHTLQEVYIELFDQIDENLKQALQKDL NLMAPGLTIQAVRVTKPKIPEAIRRNFELMEAEKTKLLIAAQKQKVVEKEA ETERKKAVIEAEKIAQVAKIRFQQKVMEKETEKRISEIEDAAFLAREKAKA DAEYYAAHKYATSNKHKLTPEYLELKKYQAIASNSKIYFGSNIPNMFVDSSC ALKYSD IRTGRESSLPSKEALEPSGENVIQNKESTG (SEQ ID NO: 346). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

This gene is expressed primarily in fetal tissue, including 8 week whole embryo, fetal liver spleen, nine week old early stage human, fetal heart, fetal liver, fetal lung, and placenta and to a lesser extent in a variety of cancers, and other normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and diseases of fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal tissues, especially the liver, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, hepatic, immune, hematopoietic, pulmonary, cardiovascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 198 as residues: Leu-68 to Lys-74, Tyr-109 to Lys-30 115, Gln-200 to Val-205, Lys-207 to Lys-214, Glu-237 to Ile-244, Ala-271 to Thr-

10

15

20

25

30

<u>- ----</u>

279, Ser-317 to Ser-329, Gln-342 to Gly-348. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of this gene (primarily fetal tissue and cancerous tissue, both of which are undergoing rapid growth) indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and disorders of fetal development. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

15

20

25

30

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:85 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1474 of SEQ ID NO:85, b is an integer of 15 to 1488, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:85, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

When tested against U937 and Jurkat cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid and T-cells, and to a lesser extent in other immune cells and tissue cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WSTGNASWEKKDNFILSADFEMMGLGNGRR SMKSPPLVLAALVACIIVLGFNYWIASSRSVDLQTRIMELEGRVRRRAAERG AVELKKNEFQGELEKQREQLDKIQSSHNFQLESVNKLYQDEKAVLVNNITTG ERLIRVLQDQLKTLQRNYGRLQQDVLQFQKNQTNLERKFSYDLSQCINQMKE VKEQCEERIEEVTKKGNEAVASRDLSENNDQRQQLQALSEPQPRLQAAGL

15

20

25

30

PHTEVPQGKGNVLGNSKSQTPAPSSEVVLDSKRQVEKEETNEIQVVNEE
PQRDRLPQEPGREQVVEDRPVGGRGFGGAGELGQTPQVQAALXVSQENPE
MEGPERDQLVIPDGQEEEQEAAGEGRNQQKLRGEDDYNMDENEAESETDKQ
AALAGNDRNIDVFNVE DQKRDTINLLDQREKRNHTL (SEQ ID NO: 347).

5 Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 9. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 9.

This gene is expressed primarily in human endometrial tumor and other tumors and to a lesser extent in a variety of other healthy adult and fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental diseases and/or disorders, particularly cancer and other proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrial tissue, cervix and uterus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 199 as residues: Asn-6 to Lys-12, Leu-65 to Phe-70, Glu-73 to His-88, Gln-123 to Gln-135, Gln-142 to Leu-156, Arg-173 to Gly-181, Asp-189 to Gln-199, Ser-204 to Arg-209, Glu-219 to Gly-225, Gly-229 to Pro-238, Ser-246 to Asn-256, Glu-263 to Arg-276. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in endometrial tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

10

15

20

25

endometrial, cervical and uterine cancer. Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:86 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3160 of SEQ ID NO:86, b is an integer of 15 to 3174, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:86, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

The translation product of this gene shares sequence homology with protein disulfide isomerase from Acanthamoeba castellanii (See Genbank Locus

ACADISPROA accession L28174, genpep locus 456013) which is thought to be important in converting proteins into their native conformations. The protein product of this gene was also shown to have homology to a phospholipase C homologue

derived from a mast cell cDNA library (See Geneseq Accession No. R99411). All references and information available through these accessions are hereby incorporated by reference herein - for example, Gene 150 (1), 175-179 (1994).

Included in this invention as preferred domains are endoplasmic reticulum targeting sequence domain and the thioredoxin family active site domain, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Proteins that permanently reside in the lumen of the endoplasmic reticulum (ER) seem to be distinguished from newly synthesized secretory proteins by the presence of the C-terminal sequence Lys-Asp-Glu-Leu (KDEL) [1,2]. While KDEL is the preferred signal in many species, variants of that signal are used by different species. This situation is described in the following table.

	Signal	Species
15	KDEL	Vertebrates, Drosophila, Caenorhabditis elegans, plants
	HDEL	Saccharomyces cerevisiae, Kluyveromyces lactis, plants
	DDEL	Kluyveromyces lactis
	ADEL	Schizosaccharomyces pombe (fission yeast)
	SDEL	Plasmodium falciparum

20

25

30

5

10

The signal is usually very strictly conserved in major ER proteins but some minor ER proteins have divergent sequences (probably because efficient retention of these proteins is not crucial to the cell). Proteins bearing the KDEL-type signal are not simply held in the ER, but are selectively retrieved from a post-ER compartment by a receptor and returned to their normal location. The concensus pattern is as follows: [KRHQSA]-[DENQ]-E-L>. Thioredoxins are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved. Bacteriophage T4 also encodes for a thioredoxin but its primary structure is not homologous to bacterial, plant and vertebrate thioredoxins. A number of eukaryotic

10

15

proteins contain domains evolutionary related to thioredoxin, all of them seem to be protein disulphide isomerases (PDI). PDI (EC 5.3.4.1) is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: - PDI major isozyme; a multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deiodinase (EC 3.8. 1.4), as glutathione-insulin transhydrogenase (EC 1.8.4.2) and as a thyroid hormone-binding protein - ERp60 (ER-60; 58 Kd microsomal protein). ERp60 was originally thought to be a phosphoinositide-specific phospholipase C isozyme and later to be a protease. - ERp72. - P5. All PDI contains two or three (ERp72) copies of the thioredoxin domain. The concensus pattern is as follows: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT]. The two C's form the redoxactive bond.

Preferred polypeptides of the invention comprise the following amino acid sequence: SLHRFVLSQAKDEL (SEQ ID NO: 348), FIKFFAPWCGHCKALAPTW (SEQ ID NO: 349), and/or FIKFYAPWCGHCKTLAPTW (SEQ ID NO: 350). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the endoplasmic reticulum 20 targeting sequence domain and thioredoxin family active site domain of the sequence referenced in Table for this gene, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of this referenced sequence. The additional contiguous amino acid residues is N-terminal or C- terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain. Alternatively, the additional contiguous amino acid residues is both N-terminal and 25 C-terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with 30 thioredoxin proteins. Such activities are known in the art, some of which are described elsewhere herein.

20

25

30

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: RRGRGVPGPRGRRRLWSAACGHCQRLQPTWN

5 DLGDKYNSMEXAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAV KYQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFELH VAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQHY ELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETGA TETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWCGHC

10 KTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLLLFRGGK KVSEHSGGRDLDS LHRFVLSQAKDEL (SEQ ID NO: 351). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in human chondrosarcoma and endothelial cells and to a lesser extent in a wide range of normal and diseased adult and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, chondrosarcoma and other cancers and proliferative disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, skeletal,

developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in chondrosarcoma, combined with the homology to protein disulfide isomerase and phospholipase C indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

10

15

20

25

30

chondrosarcoma and other cancers and proliferative disorders, and possibly as a reagent for in vitro production of proteins. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the expression in endothelial cells indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

20

25

30

related to SEQ ID NO:87 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2766 of SEQ ID NO:87, b is an integer of 15 to 2780, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:87, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed primarily in thyroid and thymus

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, thyroid diseases including thyroid cancer and diseases of function including Grave's Disease, hyper- and hypo- thyroidism as well a's Diseases of the thymus. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., endocrine, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in thyroid cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thyroid and thymus. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of

Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancrease (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothallamus, and testes. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:88 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1047 of SEQ ID NO:88, b is an integer of 15 to 1061, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:88, and where b is greater than or equal to a + 14.

20

25

15

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

The translation product of this gene shares sequence homology with collagen which is thought to be important as a structural material in a variety of human tissues and products including hair, nails, muscle and bone.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRPQGPAASPQRLRGLLLLLLQLPAPSSASEIPKGKQK AHSGRGRWWTCIMECAYKGQQECLVETGALGPMAFRVHLGSQVGMDSKEK RGNV (SEQ ID NO: 352). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle and to a lesser extent in 12 week old early stage human, epdidymus, healing groin wound, synovial hypoxia,

10

15

20

stromal cells, ulcerative colitis, breast and 8 week old embryo, as well as a variety of other normal and diseased cell types from adult and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders as well a's Diseases of smooth muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 202 as residues: Glu-32 to Glu-46, Pro-63 to Ala-71, Pro-81 to Lys-90, Ser-97 to Trp-111, Lys-130 to Ser-135, Leu-147 to Cys-154, Asp-179 to Asn-186, Ser-219 to Gly-229. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in smooth muscle and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for trearment and diagnosis of diseases of vascular diseases and/or disorders.

Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's
 Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation

of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis),

15

20

25

30

atherosclerosis, uticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, althletes foot, and ringworm).

Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chrondomalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, amd chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:89 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1328 of SEQ ID NO:89, b is an

10

15

20

25

30

E

integer of 15 to 1342, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:89, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

This gene is expressed primarily in immune cells and to a lesser extent in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T cell or B cell leukemia and various immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 203 as residues: Gly-3 to Gln-9. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system diseases such as immunodeficiencies and T cell and/or B cell leukemia. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:90 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:90, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:90, and where b is greater than or equal to a + 14.

25

5

10

15

20

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

The translation product of this gene shares sequence homology with IgE receptor. See for example, Isolation and Characterization of cDNAs coding for the Beta Subunit of the High-affinity Receptor for Immunoglobulin E, Proc. Natl. Acad. Sci. U S A. (1988 Sep.) 85(17): 6483-6487. Based on the sequence similarity, the 5 translation product of this gene is expected to share at least some biological activities with IgE receptor proteins. Such activities are known in the art, some of which are described elsewhere herein. IgE and its receptors are believed to have evolved as a mechanism to protect mammals against parasites. But other and intrinsically 10 innocuous antigens can subvert this system to provoke an allergic response. For human populations in industrialized countries, allergy and asthma now represent a far greater threat than parasitic infection, and the main impetus for current studies of the IgE system is the hope of understanding and intervening in the aetiology of allergic diseases. The high-affinity receptor for immunoglobulin (Ig) E (Fc epsilon RI) on mast cells and basophils plays a key role in IgE-mediated allergies. Fc epsilon RI is composed of one alpha, one beta, and two gamma chains, which are all required for cell surface expression of Fc epsilon RI, but only the alpha chain is involved in the binding to IgE. Fc epsilon RI-IgE interaction is highly species specific, and rodent Fc epsilon RI does not bind human IgE. New homolog can be used to develop antallergic agents. FcR deliver signals when they are aggregated at the cell surface. The aggregation of FcR having immunoreceptor tyrosine-based activation motifs (ITAMs) activates sequentially src family tyrosine kinases and syk family tyrosine kinases that connect transduced signals to common activation pathways shared with other receptors. FcR with ITAMs elicit cell activation, endocytosis, and phagocytosis. The nature of responses depends primarily on the cell type. The aggregation of FcR without ITAM does not trigger cell activation. Most of these FcR internalize their ligands, which can be endocytosed, phagocytosed, or transcytosed. The fate of internalized receptor-ligand complexes depends on defined sequences in the intracytoplasmic domain of the receptors. The coaggregation of different FcR results in positive or negative cooperation. Some FcR without ITAM use FcR with ITAM as signal transduction subunits. The coaggregation of antigen receptors or of FcR having ITAMs with FcR having immunoreceptor tyrosine-based inhibition motifs (ITIMs)

20

25

1 ._

negatively regulates cell activation. FcR therefore appear as the subunits of multichain receptors whose constitution is not predetermined and which deliver adaptative messages as a function of the environment.

The polypeptide of this gene has been determined to have four transmembrane domains at about amino acid position 51 - 67, 89 - 105, 119 - 135, and 190 - 206 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ETRVKTSLELLRTQLEPTGTVGNTIMTSQPVPN ETIIVLPSNVINFSQAEKPEPTNQGQDSLKKHLHAEIKVIGTIQILCGMMVLSL GIILASASFSPNFTQVTSTLLNSAYPFIGPFFFIISGSLSIATEKRLTKLLVHSSLV GSILSALSALVGFIILSVKQATLNPASLQCELDKNNIPTRSYVSYFYHDSLYTT DCYTAKASLAGXLSLMLICTLLEFCLAVLTAVLRWKQAYSDFPGSVLFLPH SYIGNSGMSSKMTHDCGYEELLTS (SEQ ID NO: 353). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MMVLSLGIILASASFSPNFTQVTSTLLNSAYPFIGPFFFI ISGSLSIATEKRLTKLLVHSSLVGSILSALSALVGFIILSVKQATLNPASLQC ELDKNNIPTRSYVSYFYHDSLYTTDCYTAKASLAGXLSLMLICTLLEFCL AVLTAVLRWKQAYSDFPGSVLFLPHSYIGNSGMSSKMTHDCGYEELLTS (SEQ ID NO: 354). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune system tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, immune system diseases and/or disorders such as cancer. Similarly,

10

15

20

25

30

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 204 as residues: Gln-23 to Lys-39, Glu-150 to Thr-158. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells and tissues combined with the homology to IgE receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that

10

15

20

30

influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:91 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1556 of SEQ ID NO:91, b is an integer of 15 to 1570, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:91, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 82

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GASCEGGGAAARAALGVHRSQKALLVFRRTL SNLLYMPLLRGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLLRSRLLHPEAH VPPAADRGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHGN GFPGPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSL

15

20

25

30

AFLSSLMFTSRNGLVGC GASCEGGGAAARAALGVHRSQKALLVFRRTLSNL LYMPLLRGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLLRSRLLHPEAHVPP AAD RGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHGNGFP GPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSLA FLSSLMFTSRNGLVGC (SEQ ID NO: 355). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

This gene is expressed primarily in activated T cells, and to a lesser extent in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly immunodeficeincies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 205 as residues: Pro-67 to Ser-73. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in activated T cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immunodeficiencies. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and

10

15

20

25

30

elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:92 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

30

formula of a-b, where a is any integer between 1 to 2936 of SEQ ID NO:92, b is an integer of 15 to 2950, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:92, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 83

The translation product of this gene was shown to have homology to the human transmembrane protein (See Genbank Accession No. gblAAC51364.11 (AF000959); all references and information available through this accession are hereby incorporated by reference herein; for example, Genomics 42 (2), 245-251 (1997)) which is thought to be implicated in velo-cardio-facial syndrome.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MGSAALEILGLVLCLVGWGGLILACGLPMWQVTAFLD HNIVTAQTTWKGLWMSCVVQSTGTCSAKCTTRCWL (SEQ ID NO: 356). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 22. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in dementia brain tissue, and to a lesser extent in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

15

20

25

30

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 206 as residues: Ser-201 to Tyr-217. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dementia brain tissue, combined with the homology to the transmembrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of dementia, and potentially for velo-cardio-facial syndrome. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:93 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

30

formula of a-b, where a is any integer between 1 to 1708 of SEQ ID NO:93, b is an integer of 15 to 1722, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:93, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LKRAPPGPALAKGLLQPSSTFQALETNIGDQVR RHSTAVVIREMTSYILISFVLLIGVGCIEKDQSCPVFGGRKRLHLLFVGGQLRQ VRMLRGELSCACYRPHVQALQLGGCTCF (SEQ ID NO: 357). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the adult pulmonary system.

Therefore, polynucleotides and polypeptides of the invention are useful as 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cystic fibrosis, bronchitis and any pulmonary disorders in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell 20 type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, cardiovascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary surfactant, pulmonary lavage/sputum, synovial fluid and spinal fluid) or 25 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in the pulmonary system indicates that it plays a key role in the functioning of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung and the protein product could be used either in the treatment and/or detection of these disease states. The gene

15

25

30

or gene product may also useful in the treatment and/or detection of pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:94 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 621 of SEQ ID NO:94, b is an integer of 15 to 635, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:94, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene was found to be homologous to CAM proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with CAM proteins. Such activities are known in the art, some of which are described elsewhere herein.

A preferred polypeptide varient of the invention comprises the following amino acid sequence: MLCPWRTANLGLLLILTIFLVAEAEGAAQPNNSLM LQTSKENHALASSSLCMDEKQITQNYSKVLAEVNTSWPVKMATNAVLC CPPIALRNLIIITWEIILRGQPSCTKAYKKETNETKETNCTDERITWVSRPDQ NSDLQIRTVAITHDGYYRCIMVTPDGNFHRGYHLQVLVTPEVTLFQNRNRTA VCKAVAGKPAAHISWIPEGDCATKQEYWSNGTVTVKSTCHWEVHNVSTV NCHVSHLTGNKSLYIELLPVPGAKKSSKLYIPYIILTIIILTIVGXIWLLKVNG CXKYKLNKPESTPVVEEDEMQPYAFYTEKNNPLXXTTNKVKASEALQSEV

25

30

DTDLHTL (SEQ ID NO:208). Polynucleotides encoding these polypeptides are also provided.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 271 - 287 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 288 to 348 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

This gene is expressed primarily in dendritic cells.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, autoimmunities, cancer, metastasis, wound healing, inflammation, anemias (leukemia) and other hematopoeitic disorders. Similarly, polypeptides and antibodies 15 directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell 20 types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 208 as residues: Asp-53 to Tyr-61, Pro-105 to Ile-128, Arg-133 to Leu-140, Gln-182 to Ala-188, Pro-205 to Asn-218, Gly-259 to Ala-264, Asn-290 to Ser-302, Glu-307 to Tyr-314, Tyr-317 to Lys-332. Polynucleotides encoding said polypeptides are also provided.

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from

an individual not having the disorder.

The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities,

10

15

20

25

30

immunodeficiencies (e.g. AIDS), immuno-supressive conditions (transplantation) and hematopoeitic disorders. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:95 and may have been publicly available prior to conception of

25

30

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3784 of SEQ ID NO:95, b is an integer of 15 to 3798, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:95, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VIKLICPAAFPVYFQDMARGCVCSLCASVCIFLS SLFPLLPSVHSVNIISCLLLSKCFEGLELMCEHL YQLSQLHVLHHIFSYLLCTP

15 (SEQ ID NO: 358). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in embryonic tissue and to a lesser extent in in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, cancer and neoplastic states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, differentiating, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

25

30

Ī. ---

The tissue distribution in embryonic tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental anomalies, fetal deficiencies and pre-natal disorders, as well as abnormal cell proliferation and/or differentiation, neoplastic states and cancer. Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent 15 of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, 20 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15

20

25

30

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:96 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2669 of SEQ ID NO:96, b is an integer of 15 to 2683, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:96, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

The translation product of this gene shares sequence homology with interalpha-trypsin inhibitor which is thought to be important in inhibition of trypsin and other serine proteases (See Genbank Accession No. pirlS30350lS30350; all references and information available through this accession are hereby incorporated herein by reference; for example, Eur. J. Biochem. 179 (1), 147-154 (1989), J. Biol. Chem. 264 (27), 15975-15981 (1989), and J. Biol. Chem. 266 (2), 747-751 (1991)).

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocytes, and to a lesser extent, other immune and/or hematopoietic cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the

following amino acid sequence: YXIPGSTHASGRQRGSGRGEDDSGPPPSTVINQ NETFANIIFKPTVVQQARIAQNGILGDFIIRYDVNREQSIGDIQVLNGYFVHYF APKDLPPLPKNVVFVLDSSASMVGTKLRQTKDALFTILHDLRPQDRFSIIGFS NRIKVWKDHLISVTPDSIRDGKVYIHHMSPTGGTDINGVLQRAIRLLNKYVAH SGIGDRSVSLIVFLTDG KPTVGETHTLKILNNTREAARGQVCIFTIGIGNDVD 5 FRLLEKLSLENCGLTRRVHEEEDAGSQLIGFYDEIRTPLLSDIRIDYPPSSVVQ ATKTLFPNYFNGSEIIIAGKLVDRKLDHLHVEVTASNSKKFIILKTDVPVRPQK AGKDVTGSPRPGGDGEGDXNHIERLWSYLTTKELLSSWLQSDDEPEKERLRQ RAQALAVSYRFLTPFTSMKLRGPVPRMDGLEEAHGMSAAMGPEPVVQSVR GAGTQPGPLLKKPYQPRIKISKTSVDGDPHFVVDFPLSRLTVCFNIDGQPGDIL 10 RLVSDHRDSGVTVNGELIGAPAPPNGHKKQRTYLRTITILINKPERSYLEITPS RVILDGGDRLVLPCNQSVVVGSWGLEVSVSANANVTVTIQGSIAFVILIHLYK KPAPFQRHHLGFYIANSEGLSSNCHGLLGQFLNQDARLTEDPAGPSQNLTHP LLLQVGEGPEAVLTVKGHQVPVVWKQRKIYN GEEQXDCWFARNMPPN 15 (SEQ ID NO: 359). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta and adipose tissue and to a lesser extent in several other organs and tissues including cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of developing organs and metabolic diseases, in addition to vascular diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing systems and metabolic systems, 25 expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or 30 bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 210 as residues: Lys-5 to Lys-10, Asn-33 to Lys-39, Asp-48 to Lys-54, Pro-62 to Asp-67, Asn-116 to Arg-123, His-157 to Ala-162, Val-242 to Lys-249, Val-251 to Asp-264. Polynucleotides encoding said polypeptides are also provided.

5

10

15

20

25

30

The tissue distribution in placenta, combined with the homology to interalpha-trypsin inhibitor and the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of developing and metabolic systems. This protein may play a role in the regulation of cellular division; and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease,

15

20

30

vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Polynucleotides and polypeptides of the invention are also useful for the treatment, detection, and/or prevention of inflammation, tumor invasion and metastasis, wound healing, liver disease, disseminated intravascular coagulation, alzheimer's Disease, ophthalmic disease, apoptosis, tissue remodeling, intrauterine growth retardation, preeclampsia, angiogenesis, cell migration, fetal development, trophoblast implantation, ovulation, pemphigus and psoriasis, and antiviral therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:97 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2167 of SEQ ID NO:97, b is an integer of 15 to 2181, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:97, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 88

The translation product of this gene was shown to have homology to the human colon carcinoma antigen NY-CO-7 (See Genbank and Geneseq Accession Nos. gblAAC18038.11 (AF039689) and WO9904265; all references available through this accession are hereby incorporated herein by reference; for example, Int. J. Cancer 76 (5), 652-658 (1998)).

This gene is expressed primarily in breast and breast cancer and to a lesser extent in several other organs and tissues including cancers.

15

20

25

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of reproductive organs and the gastrointestinal system, including cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, breast milk, chyme, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 211 as residues: Gly-22 to Gly-28, Leu-71 to Phe-77, Asn-101 to Val-108, Pro-122 to Ser-127, Arg-149 to Pro-154, Gly-191 to Phe-196, Pro-199 to Thr-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in breast and breast cancer tissue, combined with the homology to a colon cancer antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the reproductive systems and cancers. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell

death, as occurs in the development of some cancers, or in failure to control the extent
of cell death, as is believed to occur in acquired immunodeficiency and certain
neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

15

20

25

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:98 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1943 of SEQ ID NO:98, b is an integer of 15 to 1957, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:98, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

The translation product of this gene shares sequence homology with the amino acid and protein sequence of a Xenopus transmembrane protein of unknown function. The very 5'-end of the contig is identical to the mRNA for the human LGN mosaic protein. Based on the sequence similarity, the translation product of this gene is

15

20

25

30

expected to share at least some biological activities with LGN mosaic proteins. Such activities are known in the art, some of which are described elsewhere herein. Preferred polypeptides of the invention comprise the following amino acid sequence: PRVRPPTKALAVTFTTFVTEPLKHIGKGTGEFIKALMKEIPALLHLPVLIIMAL AILSFCYGAGKSVHVLRHIGGPEREPPQALRPRDRRRQEEIDYRPDGGAGDAD FHYRGQMGPTEQGPYAKTYEGRREILRERDVDLRFQTGNKSPEVLRAFDVPD AEAREHPTVVPSHKSPVLDTKPKETGGILGEGTPKESSTESSQSAKPVSGQDTS GNTEGSPAAEKAQLKSEAAGSPDQGSTYSPARGVAGPRGQDPVSSPCG (SEQ ID NO:339). Polynucleotides encoding such polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in small intestine and adipocytes and to a lesser extent in various other normal and transformed cell types, mostly of endocrine origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, conditions of growth and metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, gastrointestinal, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, bile, chyme, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 212 as residues: Pro-40 to Gly-68, Gly-79 to Arg-93, Phe-106 to Glu-114, Pro-122 to His-129, Thr-143 to Gly-149, Gly-155 to Ala-168,

10

15

20

25

30

Val-171 to Gly-182, Ala-195 to Pro-207, Pro-214 to Val-220. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in small intestine indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of disorders of growth and metabolism as well as endocrine abnormalities. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:99 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1098 of SEQ ID NO:99, b is an integer of 15 to 1112, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:99, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

The translation product of this gene shares sequence homology with IgE receptor beta chain which is thought to be important in immune function.

This gene is expressed primarily in kidney medulla tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and renal diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and renal systems,

expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, renal, urogenital, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney renal medulla tissue, combined with the homology to the IgE receptor beta chain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of immune and renal disorders. The protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritus, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Alternatively, this gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

10

15

20

25

30

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the

10

15

20

25

30

protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:100 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 873 of SEQ ID NO:100, b is an integer of 15 to 887, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:100, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 91

The translation product of this gene shares sequence homology with Diff 40 gene product (See Genbank Accession No. gblAAC51134.1; all references and information available through this reference are hereby incorporated herein).

Preferred polypeptides of the invention comprise the following amino acid sequence: PRVRSIKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDIN DLGTIKLSLEVTWSPFDKDDQPSAASSVNKASTVTKRFSTYSQSPPDTPS LREQAFYNMLRRQEELENGTAWSLSSESSDDSSSPQLSGTARHSPAPRPLV QQPEPLPIQVAFRRPETPSSGPLDEEGAVAPVLANGHAPYSRTLSHISEASVNA ALAEASVEAVGPKSLSWGPSPPTHPAPTHGKHPSPVPPALDPGHSATSST LGTTGSVPTSTDPAPSAHLDSVHKSTDSGPSELPGPTHTTTGSTYSAITTTHS APSPLTHTTTGSTHKPIISTLTTTGPTLNIIGPVQTTTSPTHTMPSPSSHSNSPQ YVDFCSSVCDNIFVHYVIGIFFHTLYSSKTL (SEQ ID NO:360), and/or PRVRS IKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDINDLGTIKLSLEVTWSP FDKDDQPSAASSVNKASTVTKRFSTYSQSPPDTPSLREQAFYNMLRRQEELE NGTAWSLSSESSDDSSSPQLSGTARHSPAP RPLVQQPEPLPIQVAFRRPET

15

20

25

30

PSSGPLDEEGAVAPVLANGHAPYSRTLSHISEASVNAALAEASVEAVGPKSL SWGPSPPTHPAPTHGKHPSPVPPALDPGHSATSSTLGTTGSVPTSTD (SEQ ID NO: 361). Polynucleotides encoding these polypeptides are also provided. Polypeptides of the invention do not consist of the primary amino acid sequence shown as Geneseq Accession No.W69430, which is hereby incorporated herein by reference.

This gene is expressed primarily in liver and to a lesser extent in gall bladder tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic and endocrine diseases and/or disorders, particularly hepatic and gall bladder disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hepatic, metabolic, gall bladder, gastrointestinal, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 214 as residues: Val-9 to Cys-14, Pro-42 to Thr-47, Thr-56 to Ala-64, Asp-88 to His-98, Cys-128 to Ser-136, Arg-153 to Trp-161. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in liver and gall bladder, combined with the homology to the diff 40 gene product indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study and treatment of endocrine and metabolic disorders. polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers. Representative uses are described in the "Hyperproliferative Disorders", "infectious disease", and

"Binding Activity" sections below, in Example 11, and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:101 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1234 of SEQ ID NO:101, b is an integer of 15 to 1248, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:101, and where b is greater than or equal to a + 14.

20

25

30

10

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 3 - 19 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

This gene is expressed primarily in fetal brain and to a lesser extent in pancreas tumor, melanocyte and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly neurodevelopmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are

10

15

20

25

30

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal brain tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders of the central nervous system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

20

related to SEQ ID NO:102 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1827 of SEQ ID NO:102, b is an integer of 15 to 1841, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:102, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with a probable membrane protein YGL054c -yeast (Saccharomyces cerevisiae). Moreover,

The translation product of this gene also have homology to the human and mouse cornichon protein which is known to be necessary for both anterior-posterior and dorsal-ventral pattern formation in conjunction with the EGF receptor signaling process (See Genbank Accession Nos. gblAAC98388.11 (AF104398), and splP52159; all references and information available through these accessions are hereby incorporated herein by reference; for example, Cell 81 (6), 967-978 (1995)).

The polypeptide of this gene has been determined to have two transmembrane domains at about amino acid position 57 - 73, and 121 - 137 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 - 14 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: YGCEKTTEGGRRRRRMEAVVFVFSLLDCCAL IFLSVYFIITLSDLECDYINARSCCSKLNKWVIPELIGHTIVTVLLLMSLHWF IFLLNLPVATWNIYRYIMVPSGNMGVFDPTEIHNRGQLKSHMKEAMIKLGFH LLC FFMYLYSMILALIND (SEQ ID NO:362). Polynucleotides encoding these polypeptides are also provided.

10

15

20

25

30

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in activated T-cells and to a lesser extent in endometrial tumor, T cell helper II cells, microvascular endothelial cells, Raji cells treated with cyclohexamide and umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, hematopoietic, and vascular diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 216 as residues: Ser-39 to Asn-45, Asn-103 to Ser-109. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in activated T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders involving activated T-cells. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

10

15

20

25

30

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:103 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

30

formula of a-b, where a is any integer between 1 to 671 of SEQ ID NO:103, b is an integer of 15 to 685, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:103, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARAPAPSLPPLPSPAPALAPAHSLLGLLLGRMS GSSLPSALALSLLLVSGSLLPGPGAAQNVRVQSGQDQ (SEQ ID NO: 363). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in dendritic cells and to a lesser extent in healing abdomen wound, and pancreas islet cell tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly wound healing disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 217 as residues: Gln-34 to Lys-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dendritic cells and early healing wound indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating

25

wounds to enhance the healing process. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for 10 immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatou's Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and 15 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:104 and may have been publicly available prior to conception 30 of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is

10

15

20

25

30

cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1154 of SEQ ID NO:104, b is an integer of 15 to 1168, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:104, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of aortic smooth muscle cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both smooth muscle cells, and in other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating smooth muscle cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

This gene is expressed primarily in pancreatic carcinoma, gall bladder and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic and immune diseases and/or disorders, particularly cancers, such as pancreatic carcinoma and gall bladder tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, immune, hematpoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such

10

15

20

a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 218 as residues: Lys-34 to Ile-41. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in pancreatic carcinoma and gall bladder indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancer, such as pancreatic carcinoma and gall bladder tumors.

Representative uses are described here and elsewhere herein. Alternatively, the detected calcium flux biological activity indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to miscrovascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:105 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1161 of SEQ ID NO:105, b is an integer of 15 to 1175, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:105, and where b is greater than or equal to a + 14.

25

15

20

25

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 10 - 26 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 27 to 48 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ib membrane proteins.

This gene is expressed primarily in osteosarcoma, wilm's tumor, ovarian cancer and in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammatory diseases and cancers, such as osteosarcoma, wilm's tumor and ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, renal, reproductive, immune, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 219 as residues: Ser-30 to Pro-35. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory conditions and cancer, such as osteosarcoma, wilm's tumor and ovarian cancer.

Moreover, the expression within cellular sources marked by proliferation cells.

Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases

and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

5

10

15

20

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25

30

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:106 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1007 of SEQ ID NO:106, b is an

10

15

20

25

30

integer of 15 to 1021, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:106, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSKDCVLYAFLDPGMAVPLFLYIFTLLPLLPF LLSLCFSPLTVKRSSSSESKSSL (SEQ ID NO: 364). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, ovarian, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial amniotic fluid, fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 220 as residues: Thr-28 to Ser-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in ovarian tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing cancer, e.g., ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of

10

15

20

25

30

developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:107 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

15

20

25

30

formula of a-b, where a is any integer between 1 to 816 of SEQ ID NO:107, b is an integer of 15 to 830, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:107, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

This gene is expressed primarily in macrophages and breast cancer tissue and to a lesser extent in osteoblasts and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system dysfunction; inflammation; breast cancer; cancer; osteoporosis; osteopetrosis; peristaltic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 221 as residues: Glu-16 to Ala-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. Expression in macrophages and other hematopoietic cell types indicates that this gene product is involved in the regulation of hematopoietic cell survival, proliferation, differentiation, or activation. It is involved in the control of such processes as immune surveillance, antigen presentation, T cell activation, cytokine release, and inflammation. Expression in breast cancer tissue may possibly correlate with the diagnosis and differentiation of cancerous tissue from normal breast tissue.

. 5

10

15

Expression in osteoblasts and osteoclasts may implicate this gene product in the process of bone turnover, and target it as a likely candidate for the treatment of osteoporosis and/or osteopetrosis. Finally, expressio in smooth muscle may indicate an involvement in the normal function of numerous internal organs and in the function of the digestive system.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:108 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1287 of SEQ ID NO:108, b is an integer of 15 to 1301, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:108, and where b is greater than or equal to a + 14.

		Last				360	202	30	3		43	210	919	5	7	1		
		First			Portion		7	23	3	!	42	C	4	36	9	,,	77	_
	AA First Last	AA	of			76	3	22	1	;	4	-		35	3	5	7	
L	First	AA A	of				·	-			-	-	•	-		-	•	
Ļ		SEQ	Ð	Ö	>	124	· 	125		100	071	127	į	222		128)	_
5° NT	of	First SEQ AA	AA of ID	Start Signal NO:	Pep	39		13		1.2	C	50)	313	<u> </u>	282		-
		5' NT	of		_	39		13		12	<u> </u>	50		313		282		_
	5' NT 3' NT	jo	Total Clone Clone	Seq.		1564		1757		1373		3740		1932		1196		1
	S'NT	of	Clone	Seq.		-		-		-	,	1908	***	86		-		
L	·····		Total	ZN	Seq.	1564		1757		1373		3740		1932		1196		
L	ž	SEQ	Ð	NO:	×	=		12		13		41		109		15		1
	-	······			Vector	pCMVSport	3.0	209965 Lambda ZAP	II	209965 Uni-ZAP XR		209965 Uni-ZAP XR		Uni-ZAP XR		pSport1		
		ATCC	Deposit	Nr and	Date	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98	
				cDNA	Clone ID	HDPTK41		HFXGT26		HLTGX30		HLTHG37		HLTHG37		HNTMZ90		
				Gene	No.	-		2		ж		4		4		S		

1

.

-

~

-

-_--

				Last	ΑĄ		_	O.K.	709		414		242	717		203		312	·	182
				First	AA of		Secreted	۲	30		30		19	`		61		37		22
	-	of AA First I and	Las	First SEQ AA AA	of		Sign C	,	29		29		81		!	×		36		21
	L	Ti.	2 -	$\stackrel{A}{\sim}$	Jo	3	31g	3	-				-			~	1	-	1,	_
		4		SE	f ID	<u> </u>	<u>;</u> >	_	671		95.		131		55	577		132	15	133
	S' NT	of			AA of ID	Start Signal MO.	Pen	1 6	7 0	5	176		86		4	}	200	/07		
			_	Z Z	of of		-	_		331			86		40	2	707	107	37	<u> </u>
		5' NT 3' NT	٠ -	5 	total Clone Clone	Sea.		2178	? i	1774			1674		1534		2018		2006	2
		5, N	٥٠	5		Sed.		Ŀ					47		-		255		-	
					_	Z	Seq.	2209		1774		, 60,	10/4		1534		2018		2098	
	:	N	SEO		<u> </u>	Ö	×	19		12		c	10		110		161		2	
							Vector	Uni-ZAP XR		209965 Uni-ZAP XR		nCMVCno.		3.0	pCMVSport	3.0	Uni-ZAP XR		Uni-ZAP XR	
			ATCC	Denosit	The state of the s	Nr and	Date	209965	06/11/98	209965	06/11/98	209965	20, 11, 10	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98
_					VINCE		Clone ID	HPIBX03		H6EDY30		HAMGR28			HAMGR28		HAPNZ94		HATCP77	
	<u> </u>				Gene		No.	9		7		8			∞		6		10	

-

.

		Last				146	2	122	771	Ę	7 t	5	711	155		130	601
		First	•			_ !	3	22	7	22	77	28	9	28	07	30	2
	AA First Last	AA					ì	2	i	21	i	27	 ì	27	ì	ő	ì
_	First	AA A	Jo			_		-		-		-	•	-	1	-	·
		First SEQ AA	1	Ö N	>	134		135		224	·	136		225		137	
5° NT	Jo		AA of ID	Start Signal NO:	Pep	28		86		87		62		57		205	
		5° NT	Jo		$\overline{}$	28		86	-	87		62		57		205	
	5' NT 3' NT	of	Total Clone Clone	Seq.		1746		2876		2871	· · · · ·	1052		1037		1541	
	S' NT	of	Clone	Seq.		-		-		-		30		-		-	
				NT	Seq.	1746		2876		2871		1052		1037	· <u> </u>	1541	
	Z	SEQ	Ð	ÖN	×	21		22		111		23		112		24	
					Vector	pSport1		pCMVSport	3.0	pCMVSport	3.0	pCMVSport	3.0	pCMVSport	3.0	Uni-ZAP XR	
		ATCC	Deposit	Nr and	Date	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98
				cDNA	Clone ID	HDABR72		HDPKB18		HDPKB18		HEQCCSS		HEQCCSS		HETDE26	
				Gene	No.	=		12		12		13		13		14	

Last AA of	<u> </u>					
	1 ORF 404	95	239	53	99	108
First AA of Secreted	Portion 21	31	24	22	46	21
Last AA of Sig	rep 20	30	23	21	45	20
	rep 1	-		-		-
AA SEQ ID NO:	138	139	140	141	142	143
5' NT of AA 5' NT First SEQ of AA of ID Start Signal NO:	256	253	168	233	93	148
	256	253	168	233	93	148
5° NT 3° NT of of Total Clone Clone NT Seq. Seq.	2079	1947	3354	2006	3070	2206
5' NT of Clone Seq.	-	129	-	_	-	
Total NT Seq.	2079	1947	3379	2006	3070	2227
SEQ ID NO:	25	26	27 3	28 2	 	
	╃——			 -	1 29	30
Vector	Uni-ZAP XR	209965 Uni-ZAP XR 16/11/98	Uni-ZAP XR	pCMVSport 3.0	pCMVSport 3.0	pCMVSport 2.0
ATCC Deposit Nr and Date	209965	209965	209965	209965	209965	209965 1
	НОЕДН84	HPIBT55	HSLCS05	НЪРЪЪО03	HDPDI66	НБТБQ23
Gene No.	15	16	17	18	16	20

3 1 -

-

--

Γ		100	150	AA A	Jo	ORF	7,	`		~~		166	 g	7		1	 ?		
				_	peq		f					+		\dashv	<u> </u>	-		- 6	? ——
		Firet				Portion	7	i	-	23		25	3	;	71	5	30	20	00
L	AA First Last	AA	. J.		Sig C	Pep	20	, 	5	77		24	1	18	97	ç	,	Ş	<u>,</u>
	Firs	AA	4		Sig G	rep	Ŀ			-		-		-	-	-		-	•
		SEO			;; ;	-	226		177	<u> </u>		145		22.7	/77	146	?	228	 }
5° NT	of	First SEO AA AA	of AA of ID	Q:Subjective	2	Leb	148		147	Ì		68		00		32		21	
		5' NT				COGOII	148		147	<u> </u>		68		ő	~	32	- <u></u>	21	
	5' NT 3' NT	jo	Total Clone Clone	Sec			2206		1288			3280		3300		1297		1286	
_	S' NT	jo	Clone	Sec	· 		_		1	ı		~		-	,	-		-	
				Z Z			2214		1288			3280		3300		1297		1286	<u> </u>
	Z	SEQ		Ö N	×		113	-	31			32		114		33		115	
					Vector		pCMVSport	2.0	Uni-ZAP XR			pSport1		pSport1		pCMVSport	2.0	pCMVSport	2.0
		ATCC	Deposit	Nr and	Date	2,000	506607	06/11/98	209965	06/11/98	20000	06607	06/11/98	209965	06/11/98	209965	06/11/98	209965	06/11/98
				cDNA	Clone ID	UNTUN	62001011		HE2PY40		HEONINGER	11ECIVINO		HEONM66		HKAEG43		HKAEG43)
				Gene	No.	20	04		21		23	1		22		23		23	

st ×		<u>II</u>]	1-				
	d of	412 412	272	84	153	129	. 19
First	- 01	Fortion 20	22	46	31	34	18
AA First Last SEQ AA AA ID of of	Sig	19 19	21	45	30	33	17
First AA of		2	-	-	-	-	-
AA SEC ID	<u>8</u> >	147	229	148	149	150	151
5' NT of AA First First SEQ AA AA of ID of	Signal	161	26	72	164	237	53
s, NT	Seq. Start Signal NO:	19	26	72	164	237	53
5' NT 3' NT of of Total Clone Clone	Seq.	2184	2189	949	1884	1563	1048
5° NT of Clone	Seq.	-	-	-	162	 -	
Fotal	NT Seq.	2184	2189	949	3338	1563	1048
NT SEQ ID	Ö ×	34	116 2	35 9	36 3		
0,7			R 			P 37	38
	Vector	Uni-ZAP XR	209965 Uni-ZAP XR 06/11/98	Uni-ZAP XR	Uni-ZAP XR	209965 Lambda ZAP 06/11/98 II	pSport1
ATCC Deposit	Nr and Date	209965	209965	209965	209965 1	209965 I	209965
		HLHDP65	HLHDP65	HLMDO03	HMAGK93	HMEAL02	НМКСН52
94.00	No.	24	24	25	26	27	28

_

:

		Last	AA	Jo	ORF	224		89	·	83)	99		83		161	
		First	AA of	Secreted	Portion	25		16		47		22		25		19	
	AA First Last	AA	of	Sig		24		15		46		21		24		18	
	First	AA	of	Sig	Pep	-				1		-		-		-	
	AA	SEQ		ö N	7	152		153		154		155		156		157	
5° NT	Jo	First SEQ AA	AA of ID	Start Signal NO:	Pep	188		488		130		146		120		520	
		5°NT	Jo	Start	Codon	188		488		130		146		120		520	
	5' NT 3' NT	of	Total Clone Clone	Seq.		1430		2058		2349		1559		1766		2572	
	5° NT	of	Clone	Seq.		-		209		1		-		-		427	
			Total	Z	Seq.	1430		2103		2349		1559		1766		2572	
	Z L	SEQ		ÖN.	×	39		40		41		42		43		44	
					Vector	Uni-ZAP XR		Uni-ZAP XR		pSport1		Uni-ZAP XR		Uni-ZAP XR		pCMVSport	3.0
		ATCC	Deposit	Nr and	Date	209965	06/11/98	203027	06/26/98	203027	86/97/90	203027	06/26/98	203027	06/26/98	203027	06/26/98
				cDNA	Clone ID	HCEFB69		HNFFC43		HSPMG77		HSQAC69		HSTBJ86		HLDQR62	
				Gene	No.	29		30		31		32		33		34	

:

-

:

Last	of ORF	146	142	20	64	58	- 1	3/4
First AA of	Secreted Portion	32	21	23	27	40		
AA First Last SEQ AA AA ID of of		<u></u>	20	22	26	39	2	
First of of	Pep .	-			_		-	•
AA SEC	Y 5	128	159	160	191	162	163	
S' NT of AA First of AA of ID of Start Sizes Start Start Sizes Start Sizes Start Sizes Start Sizes Start Sta	Pep	60	35	118	56	53	113	
			35	118	56	53	113	
5' NT 3' NT of of of Clone Clone NT Sea. Sea			1032	2680	1730	1275	1762	
S' NT of Clone	69		-	-	-	-	-	
	Seq. 526		1032	2680	1730	1275	1762	
NT SEQ ID NO:	× 3		94	47	48	49	50	
	Vector Uni-ZAP XR		pSport1	ZAP Express	pCMVSport	Uni-ZAP XR	Uni-ZAP XR	
ATCC Deposit Nr and	Date 203027	202027	06/26/98	203027	203027	203027		06/26/98
cDNA	Clone ID HUVDJ43	HADCP14		HBXCF95	неовитя	HL1BD22	HOEEU24	
Gene	No.	36		37	38	39	40	

		Last	AA	Jo	ORF	8		63		742		213		212		74	
		First	AA of	Secreted	Portion	22		27		27		18		27		39	
	AA First Last	AA	of	Sig	Рер	21		26		26		17		26		38	
L	First	AA	of	Sig	Pep	-		-		-		-		-		-	
	AA	SEQ	Ω	NO:	X	230		164		165		166		167		168	
S' NT	Jo	First SEQ AA	of AA of ID	Start Signal NO:	Pep	113		96		169		233		109		145	
		5° NT	of	Start	Codon	113		96		169		233		109		145	
	5' NT 3' NT	of	Clone Clone	Seq.		1763		2059		3282		1860		770		1093	
	5°NT	of	Clone	Seq.		-		-		-		-	,	-		-	·
			Total	Z	Seq.	1763		2059		3282		1860		770		1093	
	Z	SEQ	<u>a</u>	NO:	×	117		51		52		53		54		55	
					Vector	Uni-ZAP XR		Uni-ZAP XR		pCMVSport	3.0	Uni-ZAP XR		ZAP Express		Uni-ZAP XR	
		ATCC	Deposit	Nr and	Date	203027	86/97/90	203027	06/26/98	203027	06/26/98	203027	06/26/98	203027	06/26/98	203027	06/26/98
				cDNA	Clone ID	HOEEU24		HTTBR96		HWHQS55		HCEEK50		HCWBU94		HE2NR62	
				Gene	No.	40		41		42		43		44		45	

:

-

.

Γ			Last	AA	٠	10	3	47		87		7	- - -		47		\neg	232	- -	44	
			First	AA of	0.000	חברו בובת	rortion ORF	16		29		\dagger	ر ا		16 4		\neg	34		30 4	
	AA First I ast	1000	riist SEQ AA AA	Jo	5.5	31.5	3 :	15		78		5			15		-	55 —	-	29	
	Fire		AA	Jo	ij	7.5 Dan	ds.	-		_		T-	-		-		1.	-	1	-	
	AA	ָ ר	SEC.	Q	Ċ	>		109		170		12	-		172		13	C/1		174	
5° NT	of			AA of ID	Signal	Pen		167		 		199	3		153		215				
		5, NT		Jo	Start Signal NO.		201	1/7	G	×		166	· · ·		153		215		+	781	
	5' NT 3' NT	J	5	I otal Clone Clone	Seq.		632	3	26076	7007		579			13/8		1126		_ }	0/07	
	5, NT	- O	;	Clone	Seq.		F		130			153		1.	-		-		-		-
				l otal	Z	Seq.	632		7896	2		619		270	0/61		1126		2078		\dashv
	Z	SEO	_	G)	ÖN	×	56		57			- 28		9			09		61 2		\dashv
						Vector	203027 Uni-ZAP XR		pCMVSport	3.0		pBluescript		pCMVSnort	2.0	7:7	Uni-ZAP XR		203027 Uni-ZAP XR		
		ATCC	Denocit	Teoposii I	INT and	Date	203027	06/26/98	203027	06/26/98	20207	70007	06/26/98	203027			203027	06/26/98	203027	06/26/98	
				, אואקי	י אומי	Clone ID	HHSGH19		HDPGT01		HORAEII	HUNGON		HOHCA35		Т	HPMGP24	<u> </u>	HSDIE16		1
				Gene	;	No.	46		47		84	?		49		╁	 00		51		

S' NT 3' NT of o		1					S' NT					
Of of 5' NT First SEQ AA AA First Total Clone of AA of ID of of AA of NT Seq. Start Signal NO: Sig Secreted Seq. Codon Pep Y Pep Pep Por 762 1 762 433 433 175 1 16 17 1094 1 1094 173 173 176 1 24 25 1361 1 1361 112 117 1 16 17 947 1 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 33				5' NT	3, NT		Jo	AA	First	Last		
Total Clone of AA of ID of AA of NT Seq. Start Signal NO: Sig Sig Secreted Seq. Codon Pep Y Pep Pep Portion 762 1 762 433 433 175 1 16 17 1094 1 173 173 176 1 24 25 1361 1 1361 112 112 177 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	ATCC SE	1	$\overline{\diamond}$	Jo		5' NT		SEQ	AA	AA	First	Last
NT Seq. Start Signal NO: Sig Sig Sig Secreted Seq. Codon Pep Y Pep Pep Portion 762 1 762 433 433 175 1 16 17 1094 1 1094 173 173 176 1 24 25 1361 1 112 112 177 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	Deposit II	\sim		Clone	Clone		AA of	Ω	Jo	of	AA of	
Seq. Codon Pep Y Pep Pep Portion 762 1 762 433 433 175 1 16 17 1094 1 173 176 1 24 25 1361 1 172 177 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	Nr and N	Ö		Seq.	Seq.	Start	Signal	ÖN	Sig		Secreted	_
762 1 762 433 433 175 1 16 17 1094 1 1094 173 173 176 1 24 25 1361 1 1361 112 117 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	Date Vector	×				Codon	Pep	>		Pep	Portion	ORF
1094 173 173 176 1 24 25 1361 1 112 117 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	203027 Uni-ZAP XR	13		-	762	433	433	175	_	91	17	84
1094 1 173 176 1 24 25 1361 112 112 177 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	06/26/98				· · · ·				_!			
1361 1 1361 112 117 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	203027 Uni-ZAP XR 6	153	+	-	1094	173	173	176	1-	24	25	65
1361 1 1361 112 117 1 16 17 947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	06/26/98								·			
947 1 947 325 325 178 1 14 15 1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	203027 pSport1 6	4	┢	-	1361	1112	112	177	-	19	17	54
947 1 947 325 325 178 1 14 15 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	06/26/98											
1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	203027 pSport1	65	-	-	947	325	325	178	-	14	15	82
1376 1 1376 43 43 179 1 31 32 1375 1 1375 43 43 231 1 31 32	06/26/98											
1375 1 1375 43 43 231 1 31 32	203071 pCMVSport	99		-	1376	43	43	179	-	31	32	330
1375 1 1375 43 43 231 1 31 32	07/27/98 3.0				-							
	203071 pCMVSport 1	118	+	-	1375	43		231	-	31	32	71
	07/27/98 3.0	•										

:

_

-

:

Γ			Last	AA	of	ORF	2 6		1.			4		Ī		1			
-							_		_ `	<u>~</u>		54	•	_[`	<u>-</u>		147		091
		1		AA of	Secreted	Portion	30	<u>ج</u>	9	× -		23		;	67	5	67	8	07
	AA First Last		4	of	Sig		, S	7	1.	-		22		5	77	e c	07	-	7
	Firs	A	<u> </u>	Jo	Sig	Pep	<u> </u>	•		-			_	-	-	-		-	-
		SEO	}	<u>a</u>	Ö.	~	<u>8</u>		181			182		183	3	184		104	6
5° NT	Jo			AA of 1D	Start Signal NO:	Pep	548	 }	255)		309		116		213		282	
		5' NT	J.	Io	Start	Codon	548		255			309		116		213		183	
	5' NT 3' NT	of	Total Clone Clone		Seq.		2366		1086			7971		1642		921		906	
	5' N	of	ال ال		Sed.		487		-		-	-		35		-		 	,
<u>_</u>		- 0:	_	_	Z Z	Seq.	2434	-	1086		1262	1202		1642		921		906	
_	Z	SEQ	=) ;	ÖN	×	67		89	_	9	6		20		11/		72	
						Vector	Uni-ZAP XR		203071 Lambda ZAP	=	nSnort1	Popoliti		Uni-ZAP XR		Uni-ZAP XR		Uni-ZAP XR	
		ATCC	Deposit	N. o. d	ואנ מוום	Date	203071	07/27/98	203071	07/27/98	203071		07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98
				δ NO.		Clone ID	HBJIY92		HCLCW50		HDRMF68	-		HOUGG12		HEEAQ11		HEEAZ65	Ť
		<u> </u>		Gene		No.	57		28		59			09		19		62	

		Last	AA	Jo	ORF	121		162		50	-	49		46		143	
		First	AA of	Secreted	Portion	24		25		36		19	_	17		46	
	Last	AA	of	Sig	Pep	23		24		35		18		16		45	
	AA First Last	AA	of	Sig	Pep	-		-	-	-		-		-		-	
		SEQ	Q	SON SON	>	186		187		188		189		190		191	
S' NT	Jo	First SEQ AA AA	of AA of ID	Start Signal NO:	Pep	133		152		186		80		92		206	
		5' NT			Codon	133		152		186		08		92		206	· _
	5' NT 3' NT	Jo	Total Clone Clone	Seq.		089		1633		1022		1184	-	312		1370	
	5° NT	Jo	Clone	Seq.		_		-		-		-		-		38	
			Total	Z	Seq.	089		1633		1022	•	1184		312		1370	
	K	SEQ	<u> </u>	:ON	×	73		74		75		9/		77		78	
					Vector	Uni-ZAP XR	-	203071 Lambda ZAP	П	203071 pCMVSport 1		Uni-ZAP XR		pSport1		Uni-ZAP XR	
		ATCC	Deposit	Nr and	Date	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98
				cDNA	Clone ID	HEGAN94		HFXBL33		HLIBD68		HLTC033		HLYAC95		HNFGF20	
				Gene	No.	63		64	· <u> </u>	65		99		29		89	

!

-

:

		100	_	AA	Jo		1		-	4 4 ——		41	-	259		-			577	
	-	First		AA of	Secreted	Portion	27	6	5	- 17		17		28		5	2	1	7	
	AA First Last	First SEO AA AA		of	Sig	Pep	3,4	3	ç	07		16		27		2	}	-	-	_
	Firs	A A		jo	Sig			•		-		_				-		-		
		SEC		9	Ö	>	193	<u> </u>	103	}.		194		195		196	}	222	7 .	
5'NT	Jo	First		AA of ID	Start Signal NO:	Pep	125		159					\$2		341		"		
		5' NT	J		Start	Codon	125		159		177	<u> </u>		2		341		\dagger		
_	5' NT 3' NT	of	Total Clone Clone		Seq.		368		1088		1967	7001	1	1018		1984		1022	 ,	
L	5' NJ	of	71000		Seq.		-		-		-	1	-	-				78		1
			Total		Z	Seq.	368		1088		1862	700	1610	0101		2034		1022		1
	Z	SEQ	=) (: S Z	×	62		80		-∞	5	S			83		611		1
					,	Vector	203071 Uni-ZAP XR		203071 Uni-ZAP XR		203071 Uni-ZAP XR		nCMVSnort	3.0	S:	pBluescript	SK-	pBluescript	SK-	
		ATCC	Deposit	Ningad	ואו מוות	Date	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071			203071	07/27/98	203071	07/27/98	
				ANG2	7	Clone ID	HNHKS18		HSLJW78		HHFHD01		HLWAE!!		_	HCYBNSS		HCYBN55		
			·	Gene	7.7	NO.	69		70		71		72		7,3	?		73		

-

		Last	AA	Jo	ORF	869		314		348		401		324	-	8	2
		First	AA of	Secreted	Portion	24		24		24		30		31	,	91	
	First Last	AA	of	Sig		23		23		23		29	_	30	1	15	
	First	AA	of	Sig	Pep	Ŀ		-		_		-		-		-	
	AA	SEO	О	NO:	X	197		233		198		199		200		201	
5' NT	jo	First SEQ	AA of	Signal NO:	Pep	23		24		66		119	-	120		70	
		5° NT	of	Start	Codon	23		24		66	_	119		120		70	
ļ 	5' NT 3' NT	of	Total Clone Clone	Seq.		2240		2311		1488		1466		2738		1061	-
	5° NT	jo	Clone	Seq.		5		-		-		1-		2110		-	
			Total	L	Seq.	2240		2311		1488		3174		2780	-	1061	
	Z	SEQ	(l)	NO:	×	84		120		85		98		87		88	
					Vector	pSport1		pSport1		pCMVSport	3.0	pCMVSport	3.0	Uni-ZAP XR	,	Uni-ZAP XR	
		ATCC	Deposit	Nr and	Date	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98	203071	07/27/98
				cDNA	Clone ID	HEONX38		HEONX38		HLDQU79		HSYBK21		HELBC12		HTHDS25	
				Gene	No.	74		74		75		9/		77		78	

_

.

Γ		ŧ	<u></u>	⋖	<u> </u>	Ŀ	٦.	~				_		_					,	
-		_		AA	d of	ORF	_	243		93		75		340	7 70		168	•	218	
		Firet		AA of	Secreted	Portion	;	ا ا		31		31		25	ì		73		25	
\perp	AA First Last	AA	. 4		Sig	Pep	15	2		30		30	_	24			77		24	
	Firs	AA				Pep	-			_		_		-		1.	_			_
L		SEO	=	<u>.</u>	.; N	>	203	707	25	457		203		204		1	507		206	
5'NT	jo	First SEO AA	AA of 1D		Signal NO:	Pep	141		131	151		20		244			 C07		461	
_		5° NT	و ر		Start	Codon	141		131	I CT	3	2		244		263		1	 	
_	5' NT 3' NT	of	Total Clone Clone	S	Sed.		1271		1279	ì	022	?		1570	•	2014		1717	 	
<u> </u>	Š	of	Clone	Ğ					-	,	9	 }		207		78		326		
			Total			<u>.</u>	1342		1286		770	?		0/51		2950		1722		
	Z Z	SEQ	<u> </u>	Ċ	×		68		121		8	>	+	<u>-</u>		92		8		
					Vector		pSport1		pSport1		Uni-ZAP XR		203071 Hini 7 A B VB	סווו-ביאד אא		pCMVSport	3.0	Uni-ZAP XR	ı	
	({	AICC	Deposit	Nr and	Date	10000	170507	07/27/98	203071	07/27/98	203071	07/27/98	203071	11000	07/27/98	203071	07/27/98	203071	07/27/98	
				cDNA	Clone ID	UEIUOZO	0/OHIJU		HFIHO70		HPMEI86		HSOBV29			HWABY10	- 1.	HACCI17		
				Gene	Š.	92	`		62		80		81			82		83		1

-

-

		Last	AA	Jo	ORF	72		72		348		348		72		809	
		First	AA of	Secreted	Portion	25		18		27		27		30		23	
	AA First Last	AA	of	Sig	Рер	24		17	_	26		26		29		22	
	First	AA	of	Sig	Рер	-		-		_		1		1		-	
	AA	SEQ	<u>a</u>	NO N	>	235		207		208		236		209		210	
S' NT	of	First SEQ AA	of. AA of ID	Start Signal NO:	Pep	135		132		265	•	255		266		326	
		5° NT		Start	Codon	135		132		265		255		266		326	
	5' NT 3' NT	Jo	Total Clone Clone	Seq.		1380		635		3798		3793		2640		2181	
	5° NT	of	Clone	Seq.		12		-		-		1		-		-	-
				Z	Seq.	1380		635		3798		3793		2683		2181	
	Z	SEQ	Ω	Ö N	×	122		94		95		123		96		97	
					Vector	Uni-ZAP XR		203070 Uni-ZAP XR		pCMVSport	3.0	pCMVSport	3.0	pCMVSport	3.0	pCMVSport	2.0
		ATCC	Deposit	Nr and	Date	203071	07/27/98	203070	07/27/98	203070	07/27/98	203070	07/27/98	203070	07/27/98	203070	07/27/98
				cDNA	Clone ID	HACCI17		HAPQT22		HDPBO81		HDPB081		HDPGI49		HDTBV77	
				Gene	No.	83		84		85		85		.98		87	

:

-

•

:

		7	ıs	-	.	Ĺ	4 7			1		_									
-			Last	AA	1 of	OBE	-	251		225	_		20		171			47	•	139	
		Ling		AA of	Secreted	Portion		24		56			27		36			4		16	
-	AA First Last	ΑA	· · ·	ot	Sig	Pen		57		25			76		35		1	13		15	· <u> </u>
_	Firs	AA		10	Sig	Pep	1	_				1			-		-	-		-	
	AA A	SEC	=	3	Ö N	≻	1.	717		212		;	213		214		215			216	
5° NT	Jo	First SEO AA	7 4 4	01 10 PV	Start Signal NO:	Pep	24	7	$\neg \neg$	601		\top	57		20	- -	13			51	
	,	5° NT	ę		Start	Codon	24			109		23	3		20		=======================================	 -			
	5' NT 3' NT	Jo	Total Clone Clone		Seq.		1957		5	7111		887			1248		1841		- 5	949 ——	
	S' NI	Jo	Clone		seq.		-		-	-		-		1,					1		\dashv
			otal	<u> </u>		Seq.	1957		1113	7		887		+	1240	_	41		4		\dashv
	Z	SEQ	10 1	_		_	├─		╅					-			1841		289		_
 '	۷,	S	=	Ċ		\ -	86		8			8		15			102		103		
					Vector	I COLOI	pSport1		pCMVSport	3 0 %		pBluescript		nCMVSnort	3.0		Uni-ZAP XR		Uni-ZAP XR		
	C C T Y	7714	Deposit	Nr and	Date		203070	07/27/98	203070	07/27/98		203069	07/27/98	203069	07/27/98	0,000	690507	07/27/98	203069	07/27/98	
				cDNA	Clone ID	TITILITIES	HFIUE82		HHEND31			HKMND01		HLDB184		_	-		HEBEJ18		
				Gene	No.	ŝ	00		68		5	3		91		6	,		93		

.

		Last	AA	Jo	ORF	40		51		48		14		40	
		First	AA of	Secreted	Portion	30		25		30		28		15	
	Last	AA	Jo	Sig	Pep	29		24		29		27		4	
	First	AA	Jo	Sig	Pep	-		_		-		-		-	
L	of AA First Last	SEQ	Ω	ÖZ	Y	217		218		219		220		221	
5°NT	Jo	First SEQ AA AA	AA of ID	Signal	Pep	95	•	184		123		46	_	151	
		of 5' NT	Jo	Start Signal NO:	Codon	95	·	184		123		46		151	
	5' NT 3' NT	of	Clone	Seq.		1168		1175		1021		830		1301	
	5° NT	of	Total Clone Clone	Seq.		-		161		_		-		102	
		•	Total	Z	Seq.	1168		1175		1021		830		1301	
	N	SEQ	<u>(1)</u>	NO	×	104		105		106		107		108	
					Vector	203069 Lambda ZAP	II	203069 Uni-ZAP XR		203069 Uni-ZAP XR		203069 Uni-ZAP XR		203027 Uni-ZAP XR	
		ATCC	Deposit	Nr and	Date		07/27/98	203069	07/27/98	203069	07/27/98	203069	07/27/98	203027	06/26/98
				cDNA	Clone ID	HMEA148		16NDHNH		HODAE92		HODDF13		HCDCF30	_
				Gene	No.	94		95		96		26		86	

:

-

10

15

20

25

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization

15

20

25

30

probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits.

Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

10

15

20

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of

15

20

25

30

these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95%

10

15

20

25

30

"identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown inTable 1, the ORF (open reading frame), or any fragement specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence

10

15

20

25

30

that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid.

These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

5

10

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determined the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group

15

k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

20

25

30

If the subject sequence is shorter than the query sequence due to N- or Cterminal deletions, not because of internal deletions, a manual correction must be
made to the results. This is becuase the FASTDB program does not account for Nand C-terminal truncations of the subject sequence when calculating global percent
identity. For subject sequences truncated at the N- and C-termini, relative to the the
query sequence, the percent identity is corrected by calculating the number of residues
of the query sequence that are N- and C-terminal of the subject sequence, which are
not matched/aligned with a corresponding subject residue, as a percent of the total
bases of the query sequence. Whether a residue is matched/aligned is determined by
results of the FASTDB sequence alignment. This percentage is then subtracted from
the percent identity, calculated by the above FASTDB program using the specified
parameters, to arrive at a final percent identity score. This final percent identity score
is what is used for the purposes of the present invention. Only residues to the N- and

10

15

20

25

30

C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the Nterminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and Ctermini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

10

15

20

25

30

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic

10

15

20

25

30

activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of

25

30

the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, 5 where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins 15 with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. 20 Therapeutic Drug Carrier Systems 10:307-377 (1993).)

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of the present invention having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polypeptide to have an amino acid sequence which comprises the amino acid sequence of the present invention, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of the present invention or fragments thereof (e.g., the mature form and/or other fragments described herein), is

10

30

F

1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

15 Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-20 1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. 25 Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO: Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the

15

20

25

30

invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

15

20

25

30

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these

fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

5 Fusion Proteins

10

15

20

25

30

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the

10

15

20

25

30

IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral

10

15

20

25

30

vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1

10

15

20

25

30

and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

10

15

20

25

30

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or replace endogenous genetic material (e.g., coding sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with the polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter and/or enhancer) and endogenous polynucleotide sequences via homologous recombination (see, e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989); and Zijlstra et al., Nature 342:435-438 (1989), the disclosures of each of which are incorporated by reference in their entireties).

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids

15

20

30

containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or

10

15

20

25

30

translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA

be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

hybridization blocks translation of an mRNA molecule into polypeptide. Both

techniques are effective in model systems, and the information disclosed herein can

15

20

25

30

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the

10

15

20

25

30

present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as

10

15

20

25

30

deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

10

15

20

25

30

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

10

15

20

30

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic

10

15

20

25

30

anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

10

15

20

25

30

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response.

Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following 5 DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, 10 Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye 15 infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide 20 or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae,

10

Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present

invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

15

20

25

30

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vasculature (including vascular and lymphatics), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention

Chemotaxis

5

10

15

25

30

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

20 **Binding Activity**

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable

10

20

of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations,

polypeptide/molecule affixed to a solid support, chemical libraries, or natural product
mixtures. The assay may also simply comprise the steps of mixing a candidate
compound with a solution containing a polypeptide, measuring polypeptide/molecule
activity or binding, and comparing the polypeptide/molecule activity or binding to a
standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers.

The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a

candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

5

10

15

20

25

30

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, caricadic rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

10

15

20

25

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

10

15

20

25

30

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X

15

20

25

30

wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a

10

15

20

25

30

biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

20

25

30

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1

15

20

25

30

and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10

15

20

25

30

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

10

15

20

25

30

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is

10

20

defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

15 Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	•	F	
25	٠	Vector Used to Construct Library	Corresponding Deposited
		Plasmid	
30		Lambda Zap	pBluescript (pBS) pBluescript (pBS) pBK
		Uni-Zap XR	
		Zap Express	
		lafmid BA	plafmid BA
		pSport1	-
		pCMVSport 2.0	pSport1
			pCMVSport 2.0

pCMVSport 3.0

pCMVSport 3.0

pCR®2.1

pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrev Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and 10 pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to 15 the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport 1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors 20 contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday 25 Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in 30 Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional

30

plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ³²P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular 15 Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection 20 agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art. 25

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25

15

20

25

30

pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of

10

15

20

the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³²² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHybTM hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

25

30

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions: 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing

individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5

10

15

20

25

30

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1

10

20

25

30

mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., supra). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., supra).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

15 The purified protein is then renatured by dialyzing it against phosphatebuffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The

15

20

25

origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in $E \, coli$ when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Ē.i.:

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4

volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A₂₈₀ monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

25

10

10

15

20

25

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a

commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

10

15

20

25

30

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. E. coli HB101 or other suitable E. coli hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μg of a plasmid containing the polynucleotide is co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One μg of BaculoGold™ virus DNA and 5 μg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then

10

15

20

25

18

resuspended in a microcentrifuge tube containing 200 μ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 μ Ci of ³⁵S-methionine and 5 μ Ci ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include,
for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),
pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109),

20

25

30

pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing 10 cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 15 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No.209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

10

15

20

25

30

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., supra). The plasmid pSV2-neo contains a dominant selectable marker, the neo gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 -200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

30

10

15

20

25

Human IgG Fc region:

15

20

25

30

GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGC
CCAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAAA
CCCAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGT
GGTGGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGG
ACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTA
CAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACT
GGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCA
ACCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAAC
CACAGGTGTACACCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAG
GTCAGCCTGACCTGCCCCCATCCCGGGATGACCAAGACCACGCCT
CCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCCACCGTG
GACAAGAGCAGTGGCAGCAGCGGGAAACACTACAAGACCACGCCT
CCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTG
GACAAGACCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCA
TGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCCG
GTAAATGAGTGCGACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in

10

15

20

25

30

any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 μ g/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced

30

using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening 10 Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution

(1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel).

Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10⁵ cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45

minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl2 (anhyd); 0.00130 $mg/L CuSO_4$ -5 H_2O ; 0.050 mg/L of $Fe(NO_3)_3$ -9 H_2O ; 0.417 mg/L of $FeSO_4$ -7 H_2O ; 311.80 mg/L of Kcl; 28.64 mg/L of MgCl $_2$; 48.84 mg/L of MgSO $_4$; 6995.50 mg/L of 15 NaCl; 2400.0 mg/L of NaHCO $_3$; 62.50 mg/L of NaH $_2$ PO $_4$ -H $_2$ 0; 71.02 mg/L of Na₂HPO4; .4320 mg/L of ZnSO₄-7H₂O; .002 mg/L of Arachidonic Acid; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of 20 Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂0; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂0; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-25 H₂0; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalainine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tryrosine-2Na-2H₂0; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of 30 Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319

15

20

30

mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

25 Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferonsensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six

10

15

20

25

members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

			<u>JAK</u>	<u>s</u>		<u>STATS</u>	GAS(elements) or ISRE
	<u>Ligand</u>	tyk2	<u>Jak 1</u>	Jak2	Jak3		
	Traine 1						
~	IFN family						
5	IFN-a/B	+	+	. ~	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	II-10	+,	?	?	-	1,3	
	gp130 family				٠		
10	IL-6 (Pleiotrophic)	4.			2	1.2	CAC/IDEL V
	Il-11(Pleiotrophic)	+ ?	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	OnM(Pleiotrophic)		+	?	?	1,3	
٠	LIF(Pleiotrophic)	?	+	+	?	1,3	
	CNTF(Pleiotrophic)	?	+	+	?	1,3	
15	_	-/+	+	+	?	1,3	
13	G-CSF(Pleiotrophic)		+	?	?	1,3	
	IL-12(Pleiotrophic)	+	-	+	+	1,3	
	g-C family						
	T 0.4	_	+	_	+	1,3,5	GAS
20	IL-4 (lymph/myeloid)) -	+	_	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	_	+	_	+	5	GAS (IRC 1 = IFF >>Eyo)(Ign)
	IL-9 (lymphocytes)	_	+	_	+	5	GAS
	IL-13 (lymphocyte)	_	+	?	· ?	6	GAS
	IL-15	?	+	?	+	5	GAS
25		•	•	•	Т	J	UAS
•	gp140 family						,
	IL-3 (myeloid)	-	_	+	_	5	GAS (IDEISTEDS L.C)
	IL-5 (myeloid)	_		+		5	GAS (IRF1>IFP>>Ly6) GAS
	GM-CSF (myeloid)	_	-		-		
30	Cor (mycloid)		-	+	-	5	GAS
	Growth hormone fami	lv					
		<u>1¥</u> ?	_	+	_	5	
	•	•		1-	-	J	

	PRL EPO	? ?	+/- -	+	-	1,3,5	GAS(B-CAS>IRF1=IFP>>Ly6)
	Receptor Tyrosine K	<u> </u>					
5	EGF PDGF CSF-1	?	+	+	-	1,3	GAS (IRF1)
		?	+	+	-	1,3	(== 1)
		?	+	+	-	1,3	GAS (not IRF1)

15

30

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

10

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using Sall and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

15

20

25

30

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml genticin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10⁷ per transfection), and resuspend in OPTI-MEM to a final concentration of 10⁷ cells/ml. Then add 1ml of 1 x 10⁷ cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

10

15

20

25

30

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

The above protocol may be used in the generation of both transient, as well as, stable transfected cells, which would be apparent to those of skill in the art.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2x10e⁷ U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing

15

20

25

30

10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM Na₂HPO₄.7H₂O, 1 mM MgCl₂, and 675 uM CaCl₂. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting $1x10^8$ cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of $5x10^5$ cells/ml. Plate 200 ul cells per well in the 96-well plate (or $1x10^5$ cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor).

The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

- 5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)
- 5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)
- Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.
- To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker)

20 containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heatinactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and
100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is
done every three to four days. Cells are removed from the plates by scraping and
resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

10

15

20

25

30

(Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as $5x10^5$ cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to $1x10^5$ cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF-κB (Nuclear Factor κB) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF-κB regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κB appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κB is retained in the cytoplasm with I- κB (Inhibitor κB). However, upon stimulation, I- κB is phosphorylated and degraded, causing NF- κB to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κB include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF-κB promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF-κB would be useful in treating diseases. For example, inhibitors of NF-κB could be used to treat those

10

20

diseases related to the acute or chronic activation of NF-kB, such as rheumatoid arthritis.

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCC
ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGA
CTAATTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTA
TTCCAGAAGTAGTGAGGAGGCCTTTTTTTGGAGGCCTAGGCTTTTTCCAAAAAA
GCTT:3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF-KB/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF-kB/SV40/SEAP

cassette is removed from the above NF-kB/SEAP vector using restriction enzymes

SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly,

15

20

25

the NF-κB/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF-kB/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1,1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense $15 \,\mu l$ of 2.5x dilution buffer into Optiplates containing $35 \,\mu l$ of a supernatant. Seal the plates with a plastic sealer and incubate at 65° C for $30 \, \text{min}$. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)	
10	60	3	
11	65	3.25	
12	70	3.5	

13	75	2.50
14	80	3.75
15	85	4
16	90	4.25
17	95	4.5
18	100	4.75
19	105	5
20	110	5.25
21	115	5.5
22	120	5.75
23	125	6
24	130	6.25
25	135	6.5
26	140	6.75 7
27	145	7.25
28	150	7.23
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

10

15

20

25

30

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-4 (Molecular Probes, Inc.; catalog no. F-14202), used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-4 is made in 10% pluronic acid DMSO. To load the cells with fluo-4, 50 ul of 12 ug/ml fluo-4 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-4 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-4. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

10

15

5

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

20

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

25

30

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

15

20

25

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boeheringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction

20 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavadin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phospotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

10

15

20

25

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (lug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1

and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place

of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (lug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive

15

20

25

30

incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for

precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

10

15

20

25

30

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

10

15

20

25

30

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to

modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-5 permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. 10 Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. 15 USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted 20 polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both.

Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's

10

15

20

25

30

solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in

10

15

25

30

the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

20 Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

15

20

25

30

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

15

20

25

30

10

5

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide. The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The

15

20

25

30

polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely

10

15

20

25

30

differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A

15

20

25

30

time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 28: Transgenic Animals.

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e.g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i.e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40:691-698 (1994); Carver et al., Biotechnology (NY) 11:1263-1270 (1993); Wright et al., Biotechnology (NY) 9:830-834 (1991); and Hoppe et al., U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82:6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56:313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e.g., Ulmer et al., Science 259:1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and spermmediated gene transfer (Lavitrano et al., Cell 57:717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115:171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380:64-66 (1996); Wilmut et al., Nature 385:810-813 (1997)).

5

10

15

20

25

30

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89:6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265:103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse

15

20

25

30

transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 29: Knock-Out Animals.

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E.g., see Smithies et al., Nature 317:230-234 (1985); Thomas & Capecchi, Cell 51:503-512 (1987); Thompson et al., Cell 5:313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in

10

15

20

25

30

the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e.g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e.g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (i.e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e.g., lymphocytes), adipocytes. muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc. The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e.g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, <u>e.g.</u>, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U.S. Patent No. 5,399,349; and

15

20

25

Mulligan & Wilson, U.S. Patent No. 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference. Further, the hard copy of the sequence listing submitted herewith and the corresponding computer readable form are both incorporated herein by reference in their entireties.

A			
Applicant's or agent's file	D7004 DOT	International application No	
reference number	PZ031PCT	}	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications m	ade below relate to the micro	organism referred to in the description			
on page	000	lineN/A			
B. IDENTIFICATION	ONOFDEPOSIT	Further deposits are identified on an additional sheet			
Name of depositary ins	titution American Type (Culture Collection			
	,,				
Address of depositary	institution (including postar	l code and country)			
10801 University E Manassas, Virginia United States of Ar	20110-2209				
Date of deposit		Accession Number			
	July 27, 1998	203070			
C. ADDITIONAL	INDICATIONS (leave blan	k if not applicable) This information is continued on an additional sheet			
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)					
E. SEPARATE FU	RNISHING OF INDICAT	TIONS (leave blank if not applicable)			
		the International Bureau later (specify the general nature of the indications e.g., "Accession .			
	ceiving Office use only	For International Bureau use only pplication This sheet was received by the International Bureau on:			
Authorized officer Yve- PCT	te E. Simms Intermisional Division 1992)	Authorized officer			

Applicant's or agent's file			
reference number	PZ031PCT	International application No.	
			Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relatered to		
A. The indications made below relate to the mon page260	ucroorganism re , line	eterred to in the description N/A
B. IDENTIFICATIONOF DEPOSIT		
		Further deposits are identified on an additional sheet
Name of depositary institution American Ty	pe Culture Co	pllection
Address of depositary institution (including p		
	osiai coae ana co	nuntry)
10801 University Boulevard Manassas, Virginia 20110-2209		
United States of America		
Date of deposit		
July 27, 1998		Accession Number
		203069
C. ADDITIONAL INDICATIONS (leave	blank if not applica	able) This information is continued on an additional sheet
		·
D. DESIGNATED STATES FOR WHICH	INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
		
SEPADATE ETIPATION		
SEPARATE FURNISHING OF INDIC	ATIONS (leave	blank if not applicable)
lumber of Deposit")	o the Internatio	nal Bureau later (specify the general nature of the indications e.g., "Accession
		•
For receiving Office use only		
This sheet was received with the international		For International Bureau use only
With the international	application	This sheet was received by the International Bureau on:
thosis def		
uthorized officer Yvette E. Simms		Authorized officer
PCT International Divi	ston	
DOME OF THE PROPERTY OF THE PR		

A 1' I			
Applicant's or agent's file	DZOSIDOT	International application No	
reference number	PZ031PCT		Unassigned
		_ <u></u>	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

B. IDENTIFICATIONOF DEPOSIT	Parked and a second a second and a second an
	Further deposits are identified on an additional sheet
Name of depositary institution American Type Cultur	e Collection
Address of depositary institution (including postal code a	and country)
10801 University Boulevard	
Manassas, Virginia 20110-2209 United States of America	
States of America	
Date of deposit	Accession Number
July 27, 1998	203071
C. ADDITIONAL INDICATIONS (leave blank if not of	applicable) This information is continued on an additional sheet
D. DESIGNATED STATES FOR WHICH INDIC	ATIONS ARE MADE (if the indications are not for all designated States)
D. DESIGNATED STATES FOR WHICH INDIC	ATIONS ARE MADE (if the indications are not for all designated States)
D. DESIGNATED STATES FOR WHICH INDIC.	ATIONS ARE MADE (if the indications are not for all designated States)
D. DESIGNATED STATES FOR WHICH INDIC.	ATIONS ARE MADE (if the indications are not for all designated States)
D. DESIGNATED STATES FOR WHICH INDIC.	ATIONS ARE MADE (if the indications are not for all designated States)
D. DESIGNATED STATES FOR WHICH INDIC	ATIONS ARE MADE (if the indications are not for all designated States)
E. SEPARATE FURNISHING OF INDICATIONS	S(leave blank if not applicable)
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Inte	
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Inte	S(leave blank ifnot applicable)
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Inte	S(leave blank ifnot applicable)
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Inte	S(leave blank ifnot applicable)
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Inte	S(leave blank ifnot applicable)
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Intelligence of Deposit") For receiving Office use only	S(leave blank if not applicable) ernational Bureau later (specify the general nature of the indications e.g., "Accession For International Bureau use only
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Intellumber of Deposit")	S(leave blank if not applicable) Ernational Bureau later (specify the general nature of the indications e.g., "Accession For International Bureau use only
E. SEPARATE FURNISHING OF INDICATIONS The indications listed below will be submitted to the Intelligence of Deposit") For receiving Office use only	S(leave blank if not applicable) ernational Bureau later (specify the general nature of the indications e.g., "Accession For International Bureau use only

Applicant's or agent's file			
reference number	PZ031PCT	International application No.	
	. 20011 01		Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

The indications made below relate to the on page 249 LIDENTIFICATIONOF DEPOSIT Iame of depositary institution American Translates of depositary institution (including D801 University Boulevard Ianassas, Virginia 20110-2209 Inited States of America ALCOFF DEPOSIT June 26, 1998 ADDITIONAL INDICATIONS (leave DESIGNATED STATES FOR WHICE DESIGNATED STATES FOR WHIC	ype Culture Co	Further deposits are identified on an additional sheet collection Country) Accession Number 203027
ddress of depositary institution American Transcription of depositary institution (including D801 University Boulevard anassas, Virginia 20110-2209 nited States of America Description of the deposit o	postal code and co	Accession Number 203027 Cable) This information is continued on an additional sheet
ddress of depositary institution (including D801 University Boulevard anassas, Virginia 20110-2209 nited States of America ate of deposit June 26, 1998 ADDITIONAL INDICATIONS (leave	postal code and co	Accession Number 203027 Cable) This information is continued on an additional sheet
ddress of depositary institution (including D801 University Boulevard anassas, Virginia 20110-2209 nited States of America ate of deposit June 26, 1998 ADDITIONAL INDICATIONS (leave	postal code and co	Accession Number 203027 Cable) This information is continued on an additional sheet
D801 University Boulevard anassas, Virginia 20110-2209 nited States of America accordeposit June 26, 1998 ADDITIONAL INDICATIONS (leave) DESIGNATED STATES FOR WHICE	e blank if not applic	Accession Number 203027 Cable) This information is continued on an additional sheet
D801 University Boulevard anassas, Virginia 20110-2209 nited States of America accordeposit June 26, 1998 ADDITIONAL INDICATIONS (leave) DESIGNATED STATES FOR WHICE	e blank if not applic	Accession Number 203027 Cable) This information is continued on an additional sheet
D801 University Boulevard anassas, Virginia 20110-2209 nited States of America accordeposit June 26, 1998 ADDITIONAL INDICATIONS (leave) DESIGNATED STATES FOR WHICE	e blank if not applic	Accession Number 203027 Cable) This information is continued on an additional sheet
anassas, Virginia 20110-2209 nited States of America ate of deposit June 26, 1998 ADDITIONAL INDICATIONS (leave) DESIGNATED STATES FOR WHICE		203027 This information is continued on an additional sheet
nited States of America accordeposit June 26, 1998 ADDITIONAL INDICATIONS (leave) DESIGNATED STATES FOR WHICE		203027 This information is continued on an additional sheet
June 26, 1998 ADDITIONAL INDICATIONS (leave		203027 This information is continued on an additional sheet
June 26, 1998 ADDITIONAL INDICATIONS (leave		203027 This information is continued on an additional sheet
June 26, 1998 ADDITIONAL INDICATIONS (leave		203027 This information is continued on an additional sheet
ADDITIONAL INDICATIONS (leave		203027 This information is continued on an additional sheet
ADDITIONAL INDICATIONS (leave		cable) This information is continued on an additional sheet
DESIGNATED STATES FOR WHIC		
	CH INDICATIO	
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CHINDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CHINDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CHINDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CHINDICATIO	ONS ARE MADE (if the indications are not for all designated States)
	CH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
		ONS ARE MADE (1) the indications are not for all designated States)
SEPARATE FURNISHING OF INDIC		
SEPARATE FURNISHING OF INDIC		
SEPARATE FURNISHING OF INDIC		•
SEPARATE FURNISHING OF INDIC		•
SEPARATE FURNISHING OF INDIC		
OF LYNNING OF INDIC		
	CATIONS (leave	e blank if not applicable)
indications listed below will be submitted ber of Deposit")	to the Internation	onal Bureau later (specify the general nature of the indications e.g., "Accession
oci oj Deposti j		of the buildings e.g., Accession
For receive - Org		
For receiving Office use only		For International Bureau use only
This sheet was received with the internation	al application	This sheet was received by the International Bureau on:
		I L I I I I I I I I I I I I I I I I I I
	,	— Darcas On.
orized officer		Survey of the su
Vuetto F Simms		
Yvette E. Simms PCT Intermisional C		Authorized officer
C POLHEDIA MANIGICA)ivision	

Form PCT/RO/134 (July 1992)

Applicant's or agent's file		(International application N	
reference number	PZ031PCT	International application 14	Unassigned
		 	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

June 11, 1998 ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) te indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession in the submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession in the submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications e.g., "Accession in the submitted later (specify the general nature of the indications	on page	243	, line	red to in the description N/A
Accession Number June 11, 1998	B. IDENTIFICATI	ONOFDEPOSIT		Further deposits are identified on an additional sheet
DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) Per indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accessing the of Deposit") For receiving Office use only This sheet was received with the international application thorizontofficer Yvotte E. Simms	Name of depositary in	stitution American T	ype Culture Colle	
DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) Per indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accessing the of Deposit") For receiving Office use only This sheet was received with the international application thorizontofficer Yvotte E. Simms				
DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) SEPARATE FURNISHING OF INDICATIONS (teave blank if not applicable) Per indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accessing the of Deposit") For receiving Office use only This sheet was received with the international application thorizontofficer Yvotte E. Simms	Address of depositar	y institution (including	postal code and coun	ury)
Accession Number June 11, 1998 Accession Number June 11, 1998 Accession Number Z09965 ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) re indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g. "Accession number of Deporit") This sheet was received with the international application who in the international application with the international application application without as a contraction of the international Bureau on: Authorized officer Yvelie E. Simms				
June 11, 1998 Accession Number 209965 ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the international application of the indications are not for all designated States) For receiving Office use only This sheet was received with the international application Authorized officer Yvelte E. Simms	Manassas, Virginia	a 20110-2209		
June 11, 1998 ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession of Deposit") For receiving Office use only This sheet was received with the international application Authorized officer Yvette E. Simms	Onited States of A	menca		
June 11, 1998 ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession of Deposit") For receiving Office use only This sheet was received with the international application Authorized officer Yvette E. Simms	Date of deposit			Accession Number
DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) is indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession ber of Deposit") For receiving Office use only This sheet was received with the international application whorizes officer Yveltie E. Simms		June 11, 1998		
DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) This indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accessing the property of Deposit") This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvetic E. Simms	C. ADDITIONAL	INDICATIONS (lear	ve blank if not applicabl	e) This information is continued on an additional sheet
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international applicat				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international applicat				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are				
SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) se indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession and the indications e.g., "Accession and the indications e.g., "Accession and the international Bureau use only and the international application are international application are international application and the international application are	D. DESIGNATED	STATES FOR WILL	CHINDICATION	IC ADE MADE (III)
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms		JAN 125 FOR WIII	CHINDICATION	13 ARE WADE (if the indications are not for all designated States)
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms				
For receiving Office use only This sheet was received with the international application This sheet was received with the international application Authorized officer Yvette E. Simms	. SEPARATE FU	RNISHING OF IND	ICATIONS (leave b	lank if not applicable)
For receiving Office use only This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms				
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms	umber of Deposit")	ociow will be sablight	ed to the internation	at Bureau later (specify the general nature of the indications e.g., "Accession
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms				•
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms				
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms				
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms		•		
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms				
This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Yvette E. Simms	East			
thorized officer Yvette E. Simms	- -			1
Yvette E. Simms	1 has sheet was rec	-		
Yvette E. Simms		-	onal application	This sheet was received by the International Bureau on:
Yvette E. Simms PCT International Division		-	onal application	This sheet was received by the International Bureau on:
I I PCT International District 1	authorizedofficer	eived with the internati		
COL HEALTH SHOW OF I DIAMEN.	authorizedofficer	eived with the internati		

What Is Claimed Is:

- 1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
- (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
- (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
- (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
- (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
- (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
 - (f) a polynucleotide which is a variant of SEQ ID NO:X;
 - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
- (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
- (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
- 2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

- 3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
- 4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
- 5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
- 6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
- 7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
- 8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
 - 9. A recombinant host cell produced by the method of claim 8.
 - 10. The recombinant host cell of claim 9 comprising vector sequences.
- 11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

- (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
- (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
 - (g) a variant of SEQ ID NO:Y;
 - (h) an allelic variant of SEQ ID NO:Y; or
 - (i) a species homologue of the SEQ ID NO:Y.
- 12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.
- 13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.
- 14. A recombinant host cell that expresses the isolated polypeptide of claim 11.
 - 15. A method of making an isolated polypeptide comprising:
- (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and
 - (b) recovering said polypeptide.
 - 16. The polypeptide produced by claim 15.

- 17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.
- 18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:
- (a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and
- (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.
- 19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:
- (a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and
- (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.
- 20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:
 - (a) contacting the polypeptide of claim 11 with a binding partner; and
- (b) determining whether the binding partner effects an activity of the polypeptide.
 - 21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.
- 22. A method of identifying an activity in a biological assay, wherein the method comprises:
 - (a) expressing SEQ ID NO:X in a cell;
 - (b) isolating the supernatant;
 - (c) detecting an activity in a biological assay; and
 - (d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

```
<110> Human Genome Sciences, Inc.
  <120> 98 Human Secreted Proteins
  <130> PZ031.PCT
  <140> Unassigned
  <141> 1999-07-28
  <150> 60/094,657
  <151> 1998-07-30
 <150> 60/095,486
 <151> 1998-08-05
 <150> 60/095,455
 <151> 1998-08-06
 <150> 60/095,454
 <151> 1998-08-06
 <150> 60/096,319
 <151> 1998-08-12
 <160> 364
 <170> PatentIn Ver. 2.0
 <210> 1
 <211> 733
 <212> DNA
 <213> Homo sapiens
<400> 1
gggatccgga gcccaaatct tctgacaaaa ctcacacatg cccaccgtgc ccagcacctg
                                                                          60
aattegaggg tgcaccgtca gtcttcctct tccccccaaa acccaaggac accctcatga
                                                                         120
teteceggae teetgaggte acatgegtgg tggtggaegt aagecacgaa gaccetgagg
                                                                         180
tcaagttcaa ctggtacgtg gacggcgtgg aggtgcataa tgccaagaca aagccgcggg
aggagcagta caacagcacg taccgtgtgg tcagcgtcct caccgtcctg caccaggact
                                                                         240
                                                                         300
ggctgaatgg caaggagtac aagtgcaagg tetecaacaa ageceteeca acceecateg
                                                                        360
agaaaaccat ctccaaagcc aaagggcagc cccgagaacc acaggtgtac accctgcccc
                                                                         420
cateceggga tgagetgace aagaaceagg teageetgae etgeetggte aaaggettet
                                                                        480
atccaagcga catcgccgtg gagtgggaga gcaatgggca gccggagaac aactacaaga
                                                                        540
ccacgcctcc cgtgctggac tccgacggct ccttcttcct ctacagcaag ctcaccgtgg
                                                                        600
acaagagcag gtggcagcag gggaacgtct tctcatgctc cgtgatgcat gaggctctgc
                                                                        660
acaaccacta cacgcagaag agcctctccc tgtctccggg taaatgagtg cgacggccgc
                                                                        720
gactctagag gat
                                                                        733
<210> 2
<211> 5
<212> PRT
<213> Homo sapiens
<220>
<221> Site
<222> (3)
```

```
<223> Xaa equals any of the twenty naturally ocurring L-amino acids
 <400> 2
 Trp Ser Xaa Trp Ser
   1
 <210> 3
 <211> 86
 <212> DNA
 <213> Homo sapiens
 <400> 3
 gcgcctcgag atttccccga aatctagatt tccccgaaat gatttccccg aaatgatttc
                                                                           60
 cccgaaatat ctgccatctc aattag
                                                                           86
 <210> 4
 <211> 27
 <212> DNA
 <213> Homo sapiens
 <400> 4
 gcggcaagct ttttgcaaag cctaggc
                                                                           27
 <210> 5
 <211> 271
 <212> DNA
 <213> Homo sapiens
<400> 5
ctcgagattt ccccgaaatc tagatttccc cgaaatgatt tccccgaaat gatttccccg
                                                                          60
aaatatetge cateteaatt agteageaac catagteeeg cecetaacte egeceateee
                                                                         120
gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa tttttttat
                                                                         180
ttatgcagag gccgaggccg cctcggcctc tgagctattc cagaagtagt gaggaggctt
                                                                         240
ttttggaggc ctaggctttt gcaaaaagct t
                                                                         271
<210> 6
<211> 32
<212> DNA
<213> Homo sapiens
gcgctcgagg gatgacagcg atagaacccc gg
                                                                          32
<210> 7
<211> 31
<212> DNA
<213> Homo sapiens
gcgaagette gegaeteece ggateegeet e
                                                                          31
<210> 8
<211> 12
<212> DNA
```

WO 00/06698 PCT/US99/17130

3

```
<213> Homo sapiens
 <400> 8
 ggggactttc cc
                                                                         12
 <210> 9
 <211> 73
 <212> DNA
 <213> Homo sapiens
 <400> 9
 gcggcctcga ggggactttc ccggggactt tccggggact ttccatcctg
                                                                         60
 ccatctcaat tag
                                                                         73
 <210> 10
 <211> 256
 <212> DNA
 <213> Homo sapiens
<400> 10
ctcgagggga ctttcccggg gactttccg ggactttcca tctgccatct
                                                                         60
caattagtca gcaaccatag tecegeeet aacteegeee atecegeee taacteegee
                                                                        120
cagttccgcc cattctccgc cccatggctg actaattttt tttatttatg cagaggccga
                                                                        180
ggccgcctcg gcctctgagc tattccagaa gtagtgagga ggcttttttg gaggcctagg
                                                                        240
cttttgcaaa aagctt
                                                                        256
<210> 11
<211> 1564
<212> DNA
<213> Homo sapiens
<400> 11
gcggacggtg ggctgtgcaa ccttcctccc tttcttaaat gcttggggca tttgtctggc
                                                                        60
cttccctttt actgctggct gctgcctgca tctgtctctt aaccttcatt aactgtgcct
                                                                        120
atgtcaaatg gggaaccctg gtacaagata ttttcaccta tgctaaagta ttggcactga
                                                                        180
tcgcggtcat cgttgcaggc attgttagac ttggccaggg agcctctact cattttgaga
                                                                        240
attectttga gggtteatea tttgcagtgg gtgacattge cetggcaetg tacteagete
                                                                        300
tgttctccta ctcaggctgg gacaccctca actatgtcac tgaagagatc aagaatcctg
                                                                       360
agaggaacct gcccctctcc attggcatct ccatgcccat tgtcaccatc atctatatct
                                                                        420
tgaccaatgt ggcctattat actgtgctag acatgagaga catcttggcc agtgatgctg
                                                                       480
ttgctgtgac ttttgcagat cagatatttg gaatatttaa ctggataatt ccactgtcag
                                                                       540
ttgcattatc ctgttttggt ggcctcaatg cctccattgt ggctgcttct aggctttct
                                                                       600
ttgtgggctc aagagaaggc catctccctg atgccatctg catgatccat gttgagcggt
                                                                       660
tcacaccagt gccttctctg ctcttcaatg gtatcatggc attgatctac ttgtgcgtgg
                                                                       720
aagacatett ecageteatt aactaetaea getteageta etggttettt gtggggettt
                                                                       780
ctattgtggg tcagctttat ctgcgctgga aggagcctga tcgacctcgt cccctcaagc
                                                                       840
tcagcgtttt cttcccgatt gtcttctgcc tctgcaccat cttcctggtg gctgttccac
                                                                       900
tttacagtga tactatcaac tccctcatcg gcattgccat tgccctctca ggcctgccct
                                                                       960
tttacttcct catcatcaga gtgccagaac ataagcgacc gctttacctc cgaagatcgt
                                                                      1020
ggggtctgcc acaaggtacc tccaggtcct gtgtatgtca gttgctgcag aaatggattt
                                                                      1080
ggaagatgga ggagagatgc ccaagcaacg ggatcccaag tctaactaaa caccatctgg
                                                                      1140
aatcctgatg tggaaagcag gggtttctgg tctactggct agagctaagg aagttgaaaa
                                                                      1200
ggaaagctca cttctttgga ggcacctgtc cagaagcctg gcctaggcag cttcaacctt
                                                                      1260
tgaacttact ttttgaaatg aaaagtaatt tatttgtttt gctacatact gttccagact
                                                                      1320
tttaaagggg acaatgaagg tgactgtggg gaggagcatg tcaggtttgg gcttggttgt
                                                                      1380
tttagaagca cctgggtgtg cctacctact cctcttttct tttaaaaggg cccacaatgc
                                                                      1440
```

: ::

tccaatttcc tgtctccttt agagagacat gaaactatca caggtgctgg atgacaataa	1500
aagtttatgt tootaaaaaa aaaaaaaaaa aaaaaaaaaa	1560
cccg	1564
\cdot	
<210> 12	
<211> 1757	
<212> DNA	
<213> Homo sapiens	
<400> 12	
ggcacgaggt agatgggttt ttggtgtgga tgtcctttct gtttgttagt tgtccttcta	
acagacagga coctcagctg caggtctgtt ggagtaccct gcaatgtgag gtgtcagtgt	60
gcccctgctg gagggtgcct cccagttagg ctgctcgcgg gtcaggggtc agggacccac	120
cegaggagge agreegeeg treteagate tecagetgea tretaggaga accagetgeta	180
terceadage Egreagaeag ggaeatttaa gtetgeagag gtractgetg tettttest	240
tytetglyce elgececag aggiggagee tacagaggea ggeaggeete ettgaggtat	300 360
ggtgggctcc acccagttcg agcttcccgg ctgctttgtt tacctaagca agcgtggga	420
arguegggeg eccercece agreregeta coacettaca attratete agastacte	480
gerageaale agegagacte egtggggtag gaccetetga gecaggtggg ggatataatg	540
ctgtggtgtg ccattttta agcccgtcgg aaaagcgcag tattcaggtg ggagtgage	600
gatteteeg atteecagg tgccgtccgt catccette tttgattagg asagganat	660
ccctgacccc ttgcgcttcc cgagtgaggc aatgccttgc gctgctttgg ctcgtgcatg	720
gtgcgtgcac ccactgacct gcgccactg tctggcactc cctagtgaga tgaacccggt	780
acctcagatg gaaatgcaga aatcacccgt cttctgcgtt gctcaggctg ggagctgtag	840
accagagety ttectattee gecatettgg etecteece acceeege caateetaag tttatetett aacaaagtaa atactgeaca egtatttaca tgttttaaaa tattgtttat	900
ggagcactta ctatgtgttt tacggtccaa gtactttata ttattcactg actctttata	960
gcaactctat aacatattaa tagattgtct caatttcact gatgacaaaa ctggcagctt	1020
agtaactaat gagtaacaat accageceet aaaacetagg gtettteeae tacageacae	1080
telyactyae allegetette tteagacaca geactettea cetagaaaag gatgaagag	1140
cagadaging lettettet caagitigite ceatteatea atattetate aagatatata	1200 1260
gaagagttat titctaggic ticcagaagg qqctqctqac acttqqqaat tcaaaataat	1320
tacticagas ggaaatgtgg gcatagaaac tttttaaaaa ctgaattttc ccgatgtach	1380
accepted adatacagta tattattict taaaaatcag ccagaaatgc acctggagtt	1440
tractional gattiticity teccageeee attettett aagragettt tertecast	1500
titaladadg adaggatac cicccaccia taggitecci ectiatogia tattatorea	1560
gaggaradic igadatatig citgagacac aattotoatt aaggagataga cittoggag	1620
gggtgcagtg gctcacactt gtaatcccag aactttggga ggccaaggcg ggggaggatc	1680
cctgaaggcc aagagttcaa gaccagcctg gaatatatac caagatcccc tctccacaaa aaaaaaaaa aaaaaaa	1740
addada	1757
<210> 13	
<211> 1373	
<212> DNA	
<213> Homo sapiens	
<400> 13	
gtgaaatgaa atatgtatac caaaacacat aaatttaagt tttataattt tctttccttg ggatctgga aaattttctt tctgctgttt ttcattctga tagttgcttt ggctttccct	60
Attecetgee titecatatt tigaaatgit tiggicaaat tititeetgi tgagtgiga	120
reacyclicy allyctaatt ttattgacca tcatctcagt ggattatta bttgacabb	180
tctgtttcc tgttctggaa tgtttcctgc tctggaagcc atatttaac tgtatactga	240
deagette treatettaa gtatttttac aattattta chqaraatat ataactamen	300
tradicacto tagicagiga aatcaaactg ataacacaaa atgaiggigg taginaath	360
gootyadad ggatcagagg aagcactttt tagtctaagt agtcaagatc aataaagtt	420
daryayayi itigiattai itageettai gatagaagga aaaagtacag ttattanaha	480 540
gtatattaa tttatgcaaa atacacacat tatctttcca ttaacacagt agtatgacat	600
	550

```
acagttttat caccactgga ttcatgtatg cttaaaatga aaatgaaagg aacacacagg
                                                                        660
 aaagtagett ettatettee cagatacage etecaagagt geetetgetg tgtactggag
                                                                        720
 tgagtttgaa gaagttcagg taatcctctc ctaaccagca gcttaactta aagtgcatct
                                                                        780
 tcatttttgt cctaccacta aaagcttgct tattgacctc cttctttcct gtctcctaaa
                                                                        840
 tigigacigg cicicagita gactigcate tgitgetati tgiageagea attgeaggat
                                                                        900
 tcatttgctt tttctcttgc agtgtatttc tgaatgtttt ataactctgg gtatgcataa
                                                                        960
 tgtacgtata tgcatatcca tatggtcaag agttttcaga attctttgtt tgaataatga
                                                                       1020
 atgtgtgtat tccctacccc ttgttttgct ctgtcttaac atttaagctt tgcttattaa
                                                                       1080
 aatgtatttg ttttaatgct ttctctttag attgtaagct cttcatgggc agtggtcaga
                                                                       1140
 tattctgtat ttgtaacttc ccgagaatct agtattttat catgtacata atgaacatag
                                                                       1200
 aatagtattg attggaagga cagtgaagta caaaatgaga ttctaaaatg acattttcat
                                                                      1260
 tagggeteae aagtgtataa gaagtateee ttteetattt ttttgtttaa taetttagaa
                                                                       1320
 1373
<210> 14
 <211> 3740
 <212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (957)
<223> n equals a,t,g, or c
<400> 14
actetetaag gettttttgg acteaceaaa eaggettett geagtggaga tgaataeaga
                                                                        60
tcacttaagg ttgactgtgc caaatggcat aggggccctg aagctaaggg aaatggaaca
                                                                       120
ctacttctca cagggcctgt cagttcagct gtttaatgat gggtccaagg gcaaactcaa
                                                                       180
tcatttatgt ggagctgact ttgtgaaaag tcatcagaaa cctccacagg gaatggaaat
                                                                       240
taagtccaat gaaagatgct gttcttttga tggagatgca gacagaattg tttattacta
                                                                       300
ccatgatgca gatggccact ttcatctcat agatggagac aagatagcaa cgttaattag
                                                                       360
cagtttcctt aaagagctcc tggtggagat tggagaaagt ttgaatattg gtgttgtaca
                                                                       420
aactgcatat gcaaatggaa gttcaacacg gtatcttgaa gaagttatga aggtacctgt
                                                                       480
ctattgcact aagactggtg taaaacattt gcaccacaag gctcaagagt ttgacattgg
                                                                       540
agtttatttt gaagcaaatg ggcatggcac tgcactgttt agtacagctg ttgaaatgaa
                                                                       600
gataaaacaa tcagcagaac aactggaaga taagaaaaga aaagctgcta agatgcttga
                                                                       660
aaacattatt gacttgttta accaggcagc tggtgatgct atttctgaca tgctggtgat
                                                                       720
tgaagcaatc ttggctctga agggcttgac tgtacaacag tgggatgctc tctatacaga
                                                                       780
tcttccaaac agacaactta aagttcaggt tgcagacagg agagttatta gcactaccra
                                                                       840
tgctgaaaga caagcagtta caccccagg attacaggag gcaatcaatg acctggtgaa
                                                                       900
gaagtacaag ctttctcgag cttttgtccg gccctctggt acagaagatg tcgtccngag
                                                                       960
tatatgcaga agcagactca caagaaagtg cagatcacct tgcacatgaa gtgagcttgg
                                                                      1020
cagtatttca gctggctgga ggaattggag aaaggcccca accaggtttc tgaagataat
                                                                      1080
tttcatattc ctgagaaact ggacttttta caagtcttta caaaactgtc aataataatg
                                                                      1140
gcagtactaa gagatttata atcataatgt ttacaatgca gcctactgga ttgtctctag
                                                                      1200
atctgttttt cttaaacact aacagaataa ttctttataa ataggtaagc cttacacttg
                                                                      1260
ttaaagaaat ttacctctaa tttcagtctc actaatgtaa aatactggga cttaagtata
                                                                      1320
caattcagtc actaactgta cagttttatg tggggaacaa ttcatgcagg ctactggaaa
                                                                      1380
attaaatett attaccaact eettgtgata tetttgecat caccatcaca tgagcaagat
                                                                     1440
gatgttttgc agcattcccc attgctgata caaatggaga gggcagagaa gactttatac
                                                                     1500
aaccagtttt tccattgcag agtcttaaga aagattatta gatgacttac ctatatgact
                                                                     1560
aatgccatca ggaactcaga ggtatgaata gggggttgtc catccctctt ccatactgag
                                                                     1620
gtggagatgc tcatgcaata cttttaagga tgcatggtcc agccttcagt tattcttcac
                                                                     1680
tgctcttggt gaaggtatgt gggagaaaaa ctaattataa tacgtttccc agcctctgat
                                                                     1740
ggagaaggaa caccattctg ataccagaac atggttaata aggaaaagag aaaaatcccc
                                                                     1800
aaccaatctt aattgaacca agtctgaaac caatggaaaa aaaaaatggg tagtgtatat
                                                                     1860
tttgcaggtt taagacaact caggacaata aaaacaatgg actttacatg tgtatatata
                                                                     1920
tagetetett aggeaceata ateagtatga gecaacaata tttaaaettg atteaggeca
                                                                     1980
```

cattcagaca titgctctta tatacaaata titaaattaa atacaatcig aaatgigtic	2040
regulated addadaggaa adactataca acqcaqaqca qtqtqtqtqt tttaaataat	2100
tacatteaca tytaagetaa atggaaceag caatggtget caagetetta teatocotta	2160
cagadadee effectacea tetettetat tittigeetg gettigetgg aacatggfff	2220
graduated agreedatge cottattagg gaaggeattt gagtagagga taggactees	2280
tydytyteet eedcategge tigigaetii geigiigaag aetigaetga geacatigaa	2340
gudegyeagg agetgeteea taetgegeae ggtgeagatg gtgageagea agtgeegtga	2400
ciccidadee aatgitetee etgagitgie tiectetgga titticigga gaaagaaaa	2460
dycydactyg tattaataca acagacaatg tggtatgtta gaagaattaa aaatatataa	2520
accuração de la compacta de la compa	2580
dicaddacce liggigette tgagacetti tactgecatt tattagttit acatggages	2640
greedadatt gragiaatag ticccaacta gaatgegeag ataagettag traacagaaa	2700
tagctttgaa caggaataga gtcaaacata aaagttttat gttgtgcttt gtatttactc	2760
addayctee caggittetg aacceteact actgtaacca aggactaggt cacaaaatta	2820
ctatayadad daggaacada gigciitata catitcataa tatafcccct titattataa	2880
ttagttaatt cccttttatc taaatggcct aaatttgcca tgatggtagc agtgtccaaa	2940
gtgaataatt actgtcagta ctgcatcaca gagaaaggaa gggatccctc aggagacact	3000
getgettett tetgggttgt getaaacaac atagggagga aagetggace tggagtgaaa	3060
gadetydgt tagtgtgtg getetgeeat aettaeggea ceettaggea ggatataeaa	3120
aggttcctca cttataaaat gggacagtct aaaactacct tttagtagag aagtcaaatg	3180
agaaggtatg tgaaaactct gtcaactaaa tataaagact aataatttgg gtattaagag	3240
gctagtttga gaagccacct gaattacaca aacacagcta cagacatcat tctgtctaga	3300
gaaagataag agagaacagg ttggttgaac ttgggcagaa tcacagatac aattccacac	3360
taaagaatga aaataagcaa tgaactagac agaaggaaga aatcatgaag acttaggaag	3420
cagaattaca atctgtcata ttaacaaatg gagtttgcct tctaagatca gatgttgctc	3480
agaaactttc attgtttacc taataattta atatcactag tttcctagtg ggtcaagcag	3540
atgcaaaatc cagcttattt tcttctatgt gctctcaagc ttattgctta ttttaaagta	3600
aaatcctgaa aaaggaaaat attaggttgg tgcaaacgta attgcggttt ttgcattgtt	3660
gaaatttgcc gttttatatt ggagtacatt cttaaataaa tgtggttatg ttatacaaaa aaaaaaaaaa	3720
	3740
<210> 15	
<211> 1196	
<212> DNA	
<213> Homo sapiens	
2-11-11-11	
<400> 15	
ccacgcgtcc ggtgaatctg gatgatctat ttatgtcatt tatctatgtg tctcttctta	
tgccagtacc ccactgtctt gattattgtg gctttatggt gtcttgaaat taagtattgt	60
taagctttgc aatcttttcc aagattgttt tggctatagg tcctttcttt ataaagttta	120
tattaagett etecattiti etitaagaaa aaatetgetg gaatggeact gaattitag	180
gtcgatttattt gaagaattga caatattgaa cctttcaatc gatggacatg gtatgtttct	240
gcatttattt gggcctttta aaatttatct cagcaatatt ttgtagtttt agtgaagaag	300
tettgtatat atettttgtt aaatgtatte etaaatattt tgtggaaatg ttactgtaaa	360
tagtacttta atttcatttt taagtttatt getggtatac agaaatatgt tttattttg	420
taatattgac ttttatattc tgtaatctta tttaaattca cttagttcag tagtagttct	480
tatttttatt ttattgcaag tgccttaggg ataccttgtg tacacaatcc atgtcatata	540
tgaataacca acaggtttac cttttgtttc ttcttgcttt gtcgtaatgg ctaggacctt	600
ccaacctaaa tggttgaata gaagcggtgt gtgagtttat tactcttgtc ttattctcaa	660
cagcagagga tagcgtctag tttttcacta ttaaaatatg ttagctatag gtttttccg	720
tttcatgtt ttatcagatt gaagatccct tttattccta ttttgccaag agtgtcatga	780
atgratgttg aatttcacca agtacttttt cotgcatctg ttgagagaat cttgactttt	840
ctcttttatt ttgttaatgt agattcttta atgttaaact gatcttgcat tcctgatata	900
aaccatactt agtcatgata tgttatcccc tttatataat gctggattcg gtttgcaaat	960
attitigating attititiguat ctatigctcat gagacagati ggtctgtgat titicctitict	1020
tgtaagtctt gtcaggtgtc agggcttggt taattttgac acatttacaa acatatattt	1080
ttaagataaa aataaactto aaaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1140
addadadd adaaaa	1196

```
<210> 16
 <211> 2209
 <212> DNA
 <213> Homo sapiens
 <400> 16
 gagetgegee gggettggge geetggggee geegeteeee aeegtegttt teeccaeega
                                                                        60
 ggccgaggcg tcccggagtc atggccggcc tgaactgcgg ggtctctatc gcactgctag
                                                                       120
 gggttetget getgggtgeg gegegeetge egegegggge agaagetttt gågattgete
                                                                       180
 tgccacgaga aagcaacatt acagttetea taaagetggg gaccccgact etgetggcaa
                                                                       240
 aaccctgtta catcgtcatt tctaaaagac atataaccat gttgtccatc aagtctggag
                                                                       300
 aaagaatagt ctttaccttt agctgccaga gtcctgagaa tcactttgtc atagagatcc
                                                                       360
 agaaaaatat tgactgtatg tcaggcccat gtccttttgg ggaggttcag cttcagccct
                                                                       420
cgacatcgtt gttgcctacc ctcaacagaa ctttcatctg ggatgtcaaa gctcataaga
                                                                       480
gcatcggttt agagctgcag ttttccatcc ctcgcctgag gcagatcggt ccgggtgaga
                                                                       540
gctgcccaga cggagtcact cactccatca gcggccgaat cgatgccacc gtggtcagga
                                                                       600
ttggaacctt ctgcagcaat ggcactgtgt cccggatcaa gatgcawgaa ggagtgaaaa
                                                                       660
tggccttaca cctcccatgg ttccacccca gaaatgtctc cggcttcagc attgcaaacc
                                                                       720
gctcatctat aaaacgtctg tgcatcatcg agtctgtgtt tgagggtgaa ggctcagcaa
                                                                       780
ccctgatgtc tgccaactac ccagaaggct tccctgagga tgagctcatg acgtggcagt
                                                                       840
ttgtcgttcc tgcacacctg cgggccagcg tctccttcct caacttcaac ctctccaact
                                                                       900
gtragaggaa ggaggagcgg gttgaatact acatcccggg ctccaccacc aaccccgagg
                                                                       960
tgttcaaget ggaggacaag cageetggga acatggeggg gaacttcaac etetetetge
                                                                      1020
aaggetgtga ccaagatgee caaagteeag ggateeteeg getgeagtte caagttttgg
                                                                      1080
tccaacatcc acaaaatgaa agcaataaaa tctacgtggt tgacttgagt aatgagcgag
                                                                      1140
ccatgtcact caccatcgag ccacggcccg tcaaacagag ccgcaagttt gtccctggct
                                                                      1200
gtttcgtgtg tctagaatct cggacctgca gtagcaacct caccctgaca tctggctcca
                                                                      1260
aacacaaaat ctccttcctt tgtgatgatc tgacacgtct gtggatgaat gtggaaaaam
                                                                      1320
ccataagytg cacagaccac cggtactgcc aaaggaaatc ctactcacty caggtgccca
                                                                      1380
gtgacatect yeametgeet gtggagetge atgaettete etggaagetg etggtgeeca
                                                                      1440
aggacagget cagectggtg ctggtgccag cccagaaget gcagcagcat acacacgaga
                                                                      1500
agecetgeaa caccagette agetaceteg tggecagtge catacecage caggacetgt
                                                                      1560
actteggete ettetgeeeg ggaggeteta teaageagat eeaggtgaag eagaacatet
                                                                      1620
cggtgaccct tcgcaccttt gcccccagct tccgacaaga ggcctccagg cagggtctga
                                                                      1680
cggtgtcctt tataccttat ttcaaagagg aaggcgtttt cacggtgacc cctgacacaa
                                                                      1740
aaagcaaggt ctacctgagg acccccaact gggaccgggg cctgccatcc ctcacctctg
                                                                      1800
tgtcctggaa catcagcgtg cccagagacc aggtggcctg cctgactttc tttaaggagc
                                                                      1860
ggagcggcgt ggtctgccag acagggcgcg cattcatgat catccaggag cagcggaccc
                                                                      1920
1980
acagettetg ggteaacate tytaaytgma geeceacgag eggeaageag etagaeetge
                                                                      2040
tetteteggt gacaettace ceaaggactg tggaettgae tgteatecte ategeagegg
                                                                      2100
tgggaggtgg agtettaetg etgtetgeee tegggeteat eatttgetgt gtgaaaaaaa
                                                                      2160
aaaaaarama aacaaggggc cccgctgtgg gtatctacaa tggcaacat
                                                                     2209
<210> 17
<211> 1774
<212> DNA
<213> Homo sapiens
<400> 17
taggcacccc aggettttac actttatgtt ceggttegta tgttgtgtga attgtaaccg
                                                                       60
gataccaatt tcacacaagg aaccagetta tgaccatgat tacgecaage tegaaattaa
                                                                      120
ccctcactaa agggaacaaa agctggagct ccaccgcggt ggcggccgct ctagaactag
                                                                      180
tggatccccc gggctgcagg aattcggcac gagcggctgc gggcgcgagg tgaggggcgc
                                                                      240
gaggttccca gcaggatgcc ccggctctgc aggaagctga agtgagaggc ccggagaggg
                                                                      300
cccagcccgc ccggggcagg atgaccaagg cccggctgtt ccggctgtgg ctggtgctgg
                                                                      360
ggtcggtgtt catgatcctg ctgatcatcg tgtactggga cagcgcaggc gccgcgcact
                                                                      420
```

totacttgca cacgtootto totaggoogo acacggggco geogotgcoc acgcocgggc

```
cggacaggga cagggagete aeggeegaet eegatgtega egagtttetg gacaagttte
                                                                         540
 tcagtgctgg cgtgaagcag agtgacettc ccagaaagga gacggagcag ccgcctgcgc
                                                                         600
 cggggagcat ggaggagaac gtgagaggct acgactggtc cccgcgcgac gcccggcgca
                                                                         660
 gcccagacca gggccggcag caggcggagc ggaggagcgt gctgcggggc ttctgcgcca
                                                                         720
 actccageet ggeetteece accaaggage gegeattega egacateece aacteggage
                                                                         780
 tgagccacct gatcgtggac gaccggcacg gggccatcta ctgctacgtg cccaaggtgg
                                                                         840
 cctgcaccaa ctggaagcgc gtgatgatcg tgctgagcgg aagcctgctg caccgcggtg
                                                                         900
 cgccctaccg cgacccgctg cgcatcccgc gcgagcacgt gcacaacgcc agcgcgcacc
                                                                         960
 tgaccttcaa caagttctgg cgccgctacg ggaagctctc ccgccacctc atgaaggtca
                                                                        1020
 ageteaagaa gtacaccaag tteetetteg tgegegaeee ettegtgege etgateteeg
                                                                        1080
 ccttccgcag caagttcgag ctggagaacg aggagttcta ccgcaagttc gccgtgccca
                                                                        1140
 tgctgcggct gtacgccaac cacaccagcc tgcccgcctc ggcgcgcgag gccttccgcg
                                                                        1200
 ctggcctcaa ggtgtccttc gccaacttca tccagtacct gctggacccg cacacggaga
                                                                        1260
 agetggegee etteaacgag caetggegge aggtgtaceg eetetgeeac eegtgeeaga
                                                                        1320
 tcgactacga cttcgtgggg aagctggaga ctctggacga ggacgccgcg cagctgctgc
                                                                        1380
 agctactcca ggtggaccgg cagctccgct tccccccgag ctaccggaac aggaccgcca
                                                                        1440
 gcagctggga ggaggactgg ttcgccaaga tccccctggc ctggaggcag cagctgtata
                                                                        1500
aactctacga ggccgacttt gttctcttcg gctaccccaa gcccgaaaac ctcctccgag
                                                                        1560
actgaaagct ttcgcgttgc tttttctcgc gtgcctggaa cctgacgcac gcgcactcca
                                                                        1620
gtttttttat gacctacgat tttgcaatct gggcttcttg ttcactccac tgcctctatc
                                                                        1680
cattgagtac tgtatcgata ttgtttttta agattaatat atttcaggta tttaatacga
                                                                        1740
aaaaaaaaa aaaaaaactc gagggggggc ccgg
                                                                        1774
<210> 18
<211> 1674
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (1649)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1663)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1665)
<223> n equals a,t,g, or c
<400> 18
caagttggta cgcctgcagg taccggtccg gaattcccgg gtcgacccac gcgtccggtc
                                                                         60
gaagataggt tcgagacggt ggatgttgca gctgatcatg cagttgggtt cggtgctgct
                                                                        120
cacacgctgc cccttttggg gctgcttcag ccagctcatg ctgtacgctg agagggctga
                                                                        180
ggcacgccgg aagcccgaca tcccagtgcc ttacctgtat ttcgacatgg gggcagccgt
                                                                        240
gctgtgcgct agtttcatgt cctttggcgt gaagcggcgc tggttcgcgc tgggggccgc
                                                                        300
actccaattg gccattagca cctacgccgc ctacatcggg ggctacgtcc actacgggga
                                                                        360
ctggctgaag gtccgtatgt actcgcgcac agttgccatc atcggcggct ttcttgtgtt
                                                                        420
ggccagcggt gctggggagc tgtaccgccg gaaacctcgc agccgctccc tgcagtccac
                                                                        480
cggccaggtg ttcctgggta tctacctcat ctgtgtggcc tactcactgc agcacagcaa
                                                                        540
ggaggaccgg ctggcgtatc tgaaccatct cccaggaggg gagctgatga tccagctgtt
                                                                        600
cttcgtgctg tatggcatcc tggccctggc ctttctgtca ggctactacg tgaccctcgc
                                                                        660
tgcccagate etggetgtae tgctgcccce tgtcatgctg etcattgatg gcaatgttge
                                                                        720
ttactggcac aacacgcggc gtgttgagtt ctggaaccag atgaagctcc ttggagagag
                                                                        780
```

```
tgtgggcatc ttcggaactg ctgtcatcct ggccactgat ggctgagttt tatggcaaga
                                                                          840
 ggctgagatg ggcacaggga gccactgagg gtcaccctgc cttcctcctt gctggcccag
                                                                          900
 ctgctgttta tttatgcttt ttggtctgtt tgtttgatct tttgcttttt taaaattgtt
                                                                          960
 ttttgcagtt aagaggcagc tcatttgtcc aaatttctgg gctcagcgct tgggagggca
                                                                         1020
 ggagccctgg cactaatgct gtacaggttt ttttcctgtt aggagagctg aggccagctg
                                                                         1080
 cccactgagt ctcctgtccc tgagaaggga gtatggcagg gctgggatgc ggctactgag
                                                                         1140
 agtgggagag tgggagacag aggaaggaag atggagattg gaagtgagca aatgtgaaaa
                                                                         1200
 attcctcttt gaacctggca gatgcagcta ggctctgcag tgctgtttgg agactgtgag
                                                                         1260
 agggagtgtg tgtgttgaca catgtggatc aggcccagga agggcacagg ggctgagcac
                                                                         1320
 tacagaagtc acatgggttc tcagggtatg ccaggggcag aaacagtacc ggctctctgt
                                                                         1380
 cactcacctt gagagtagag cagaccctgt tctgctctgg gctgtgaagg ggtggagcag
                                                                        1440
 gcagtggcca gctttgccct tcctgctgtc tctgtttcta gctccatggt tggcctggtg
                                                                        1500
 ggggtggagt teceteceaa acaccagace acacagteet ecaaaaataa acattttata
                                                                        1560
 tagamaaaaa aaaaaaaaaa aagggcggcc gctctagagg atccctcgag gggcccaagc
                                                                        1620
 ttacgcgtgc atgcgacgtc atagctctnt ccctatagaa gtngnaaagg gttc
                                                                        1674
 <210> 19
 <211> 2018
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (2010)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (2012)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (2014)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (2016)
<223> n equals a,t,g, or c
cagcaatgaa atcctgcttt cttttcctca gaactactat attcagtggc taaatggctc
                                                                         60
cctgattcat ggtttgtgga atcttgcttc ccttttttcc aacctttgyt tatttgtatt
                                                                        120
gatgcccttt gcctttttct ttctggaatc agaaggcttt gctggcctga aaaagggaat
                                                                        180
ecgagecege attttagaga etttggteat gettettett ettgegttae teattettgg
                                                                        240
gatagtgtgg gtagetteag caeteattga caacgatgee geaageatgg aatetttata
                                                                        300
tgatctctgg gagttctatc taccctattt atattcctgt atatcattga tgggatgttt
                                                                        360
gttacttctc ttgtgtacac cagttggcct ttctcgtatg ttcacagtga tgggtcactt
                                                                        420
gctagtgaag ccaacaattc ttgaagacct ggatgaacaa atttatatca ttaccttaga
                                                                        480
ggaagaagca ctccagagac gactaaatgg gctgtcttca tcggtggaat acaacataat
                                                                        540
ggagttggaa caagaacttg aaaatgtaaa gactcttaag acaaaattag agaggcgaaa
                                                                        600
aaaggettea geatgggaaa gaaatttggt gtateeeget gttatggtte teettettat
                                                                        660
tgagacatcc atctcggtcc tcttggtggc ttgtaatatt ctttgcctat tggttgatga
                                                                        720
aacagcaatg ccaaaaggaa caagggggcc tggaatagga aatgcctctc tttctacgtt
                                                                        780
tggttttgtg ggagctgcgc ttgaaatcat tttgattttc tatcttatgg tgtcctctgt
                                                                        840
tgtcggcttc tatagccttc gattttttgg aaactttact cccaagaaag atgacacaac
                                                                        900
tatgacaaag atcattggaa attgtgtgtc catcttggtt ttgagctctg ctctgcctgt
```

```
gatgtcgaga acactgggaa tcactagatt tgatctactt ggcgactttg gaaggtttaa
                                                                         1020
  ttggctggga aatttctata ttgtattatc ctacaatttg ctttttgcta ttgtgacaac
                                                                         1080
 attgtgtctg gtccgaaaat tcacctctgc agttcgagaa gaacttttca aggccctagg
                                                                         1140
 gcttcataaa cttcacttac caaatacttc aagggattca gaaacagcca agccttctgt
                                                                         1200
 aaatgggcat cagaaagcac tgtgagacgc acagacggcg tcttctgcca ccaagagacc
                                                                         1260
 cgagaactcc agattcacga cattcctgtc ccatgtagaa gcatttccat tcaaccgtgg
                                                                         1320
 cccctcttca gaacctagac ctatcagtgc cattttttt tcataatcta cgaagaactt
                                                                         1380
 ggctatggct gatctttttt aaatttaact ttctgatgga ccctgtagtt tccagttaag
                                                                         1440
 tgcagattcc ttacagacat atagaacagc gcattcttct gtagacattt gctcatgttg
                                                                         1500
 gtaaatacaa tcacccatat gaaaaaattg ttttcacctg atatgaaaat gttagaaaag
                                                                        1560
 gcaaactccg ggacttctaa agatttactt aaatcccatt atgtacttta ttcagaatgt
                                                                        1620
 agaagctgac ttgaaaggca tccttggtac taagtgaagc ttattcagaa aatgcatttt
                                                                        1680
 tcaaatgcaa tggcaactgc ttgtagatat catttttgca gtgtatgttg gagctgtaat
                                                                        1740
 ggttgcaatt atgtttetta ttteettaaa agcaaaaage gtagtttetg atttatgtta
                                                                        1800
 tagaatgata ctgattagac tttgagccaa ggggaaaata ctaaattctt ttaaacctgg
                                                                        1860
 agcettagag agceaeagga atatettetg ttgtaeagte taataagetg tggtaggaag
                                                                        1920
 tatcatgtaa tcacagttta atgacagttt atgtatatat ataattcagt attccctcga
                                                                        1980
 gggggggccc ggtacccaat tcgcccaagn tnancnag
                                                                        2018
 <210> 20
 <211> 2098
 <212> DNA
 <213> Homo sapiens
 <400> 20
 ggcacgaggg accgagctat tctcctggga ctggctatga tggtgtgctc catcatgatg
                                                                          60
 tattttctgc tgggaatcac actcctgcgc tcatacatgc agagcgtgtg gaccgaagag
                                                                         120
 teteaatgea eettgetgaa tgegteeate aeggaaaeat ttaattgete etteagetgt
                                                                         180
 ggtccagact gctggaaact ttctcagtac ccctgcctcc aggtgtacgt taacctgact
                                                                         240
 tetteegggg aaaageteet eetetaceae acagaagaga caataaaaat caateagaag
                                                                         300
 tgctcctata tacctaaatg tggaaaaaat tttgaagaat ccatgtccct ggtgaatgtt
                                                                         360
 gtcatggaaa acttcaggaa gtatcaacac ttctcctgct attctgaccc agaaggaaac
                                                                         420
cagaagagtg ttatcctaac aaaactctac agttccaacg tgctgttcca ttcactcttc
                                                                         480
 tggccaacct gtatgatggc tgggggtgtg gcaattgttg ccatggtgaa acttacacag
                                                                         540
tacctctccc tactatgtga gaggatccaa cggatcaata gataaatgca aaaatggata
                                                                         600
aaataatttt tgttaaagct caaatactgt tttctttcat tcttcaccaa agaaccttaa
                                                                         660
gtttgtaacg tgcagtctgt tatgagttcc ctaatatatt cttatatgta gagcaataat
                                                                         720
gcaaaagctg ttctatatgc aaacatgatg tctttattat tcaggagaat aaataactgt
                                                                        780
tttgtgttgg ttggtggttt tcataatctt atttctgtac tggaactagt actttcttct
                                                                        840
ctcattccgc caaaacaggg ctcagttatt catttgccaa gcttcgtgga ggaatgtagg
                                                                        900
tgacatcaat gtgataaagt ctgtgttctg agttgtcaga tctcttgaag acaatatttt
                                                                        960
tcatcactta ttgtttacta aagctacagc caaaaatatt tttttttctt attctaaact
                                                                       1020
gagccctata gcaagtgaag ggaccagatt tcctaattaa aggaagttag gtactttct
                                                                       1080
tgtatttttt accatatcac tgtaaagaag aggggaaacc cagccagcta cttttttca
                                                                       1140
tcacttttta ttcataactt cagatttgta aaactaattt ccaaaatata agctgttttc
                                                                       1200
attagccagt tctataatat cttcctgtga tttatgtaga aaatgaacac accccttttc
                                                                       1260
catttaagac cctgctactg tgtgaagaga tgatacttac aaggagtgtc attacctgtg
                                                                       1320
agetgaetga atgttggtag gtgetecatt acaatecagg aaagtetgtg ttaetgatat
                                                                       1380
ttgtgtggaa atctttattt cacttcaatt taaccattag atggtaaaat taagatgcta
                                                                       1440
cttgttggta aaaattggtg gactggtttc aatgggtaaa tgtgttgtgg caaattaatg,
                                                                       1500
tgttggaata ttgctctttg tgaatttgtg cttaagtcaa tgaatgtgta gtatctcctt
                                                                       1560
ctgacaagca ttccctattg ggattttaaa gctatgtgca cagaatatta gtctcttcta
                                                                       1620
catgittiat tittctatti ataattccct tittigtigt tatatittat acacagaata
                                                                       1680
gatctttttt ctaacacata tttgaactga ataacagact taaagaaagc ctttgttcac
                                                                       1740
attgctattt acttttgtgt ttgggggaaa atacgaggga ttgattttaa ataaaaaaca
                                                                       1800
ttccatcttt catttaatat caatatcaaa agaagaagac aaacatctat ctttctcatc
                                                                       1860
tatatttaag tacctttttg taatgtagta tcaaagtttt ttaggtaatg caaaatttta
                                                                       1920
caaatcattt gtggaatgaa tggtaaaact aatctgatga aatggaaaat tattctgcaa
```

tattgtaatt ttcaaatgtg	catagtttga taattaaatt	cttttcataa tttttaaaaa	a gcaaataaa a aaatctaaa	t ccctagga a aaaaaaaa	tg taatcaggac aa aaaaaaaa	2040 2098
<210> 21				•		
<211> 1746						
<212> DNA						
<213> Homo	sapiens					
<400> 21						
gegetegege	acctccagct	cgggccgatg	tggaagctt	t ggagagctg	a agagggcgcg	60
geggegeteg	geggeget	cttcctgctg	ctcttcgcg	c taggggtcc	a ccaactacta	120
atctattccc	taggaggeta	cttcccccg	gggccgccg	g ggctgccat	t tatcggcaac	180
gtgtacggag	aggtagette	accegagett	ccccatgtc	t acatgagaa	a gcagagccag	240
gactaaccaa	agcetteeta	attagaatta	tagaggagg	g agggccgcc	a ggctggcccg	300
teteeqtqtq	Coactaccc	tttgagggac	accoactto	: cgggccgac	t agtcgggacc a agggagaagt	360
cctcgacgcc	gtgcccctt	acadadadaa	ccccccccc	cyggicigg	a agggagaagt c cactccgggc	420
cgaggctccg	aggcgatcca	gtcctgattr	tcccgctac	gccggtgac	c ttgctcctgc	480
gcctgcgccg	tttggctcgc	caqccqcqcc	gccacttcac	gecegage	g gacgcatgc	540
ctcaggtgcg	ggcgtcttgc	gagtcggcct	cacaactete	, geceaggge	c acgcggcttg	600
ccggaaaatt ,	aaygcgttct	gagttctaga	tggttaatac	r caggttett	gatatataca	660 720
geegaegaac	gactggtgta	ggcgtttgct	gtgagaatgc	i agaatgcag	T GGSSCGCCC	780
eguccyayaa	acaaaccca	ggaaacgatt	gtgaacgcgt	gaatgaatt	atgactasaa	840
cccgccgcgg g	gggtcctaca	gcgcagatgg	taatgccgtt	ctgactggci	gggaacggca	900
ccctagcaga	tacttaaaag	gcgccttctg	tataccacto	teactgccaa	cttaataact	960
catttaaaac	cataaccag	ccggtgaggt	cagtacttca	ctootootoa	ttctgcggag	1020
gggaaagcag (cacggaaatg	ccctgtgact	ggcagcggaa	aaggcgacca	ccacttatat	1080
ctattttatt	egacgtccgg .	agggggcagg	agtttccacg	ggtcctggga	cagageteae	1140
gaaggcagat :	coggettt	cttatttata	tgcaactaca	ggcctgacgo	tagcggtgaa	1200
attategge a	acageetete (aaggagttgg	cagatgagtg	ggagagagaa	aactaatctc	1260
attatcggcc a	aggtttgac	cttaaatoot	tgaaggaaaa	gtacagggat	gtttggcaac	1320
tgtggtattt c	acacaatac	cacacctat	aatcccacca	agtttttaca	aggattggtc	1380
cggatcacga a	atcaggaga i	Cgagaccgt	cataactaac	accortonance	ccgaggcggg	1440
Ctaaaagaat a	caaaaaatt g	ggcgggggt	aataacaaac	acctgtggt	ccacctatta	1500
aaaaccaa a	gcaggagag t	ggcgtgaac	ccqqqaqqca	gagettteag	tgagccgaga	1560 1620
eegegeeact g	cactccage o	tgggcaaca	gagccagact	ccgtctcaaa	aaaaaaaaa	1680
uuuaaagggcg g	ccgctctag a	aggatccaag	cttacgtacg	cgtgcatgcg	acgtcaatag	1740
ctcttc						1746
	•					
<210> 22						
<211> 2876						
<212> DNA						
<213> Homo s	apiens					
<400> 22						
ccacgcgtcc g	gcggcgagg c	ggcggtggt g	gctgagtcc	gtggtggcag	aggcgaaggc	60
gacageteta g	yygriggea c	cggccccga c	aggaggatg	caaatccaaa	taggggtgag	120
gergergerg t	grgeggtge t	gctgagett d	racctcaaca	tecteggatg	220220022	180
ccaggatgaa t	ccttagatt c	caagactac t	ttgacatca	gatgagtcag	taaaggagga	240
caccaccyca g	ycagagtag t	tgctggtca a	atatttett	gattcagaag	aatctcaatt	300
agaateetee at	cccaagaag a	ggaagacag c	ctcaagagc	Caagaggggg	aaagtgtgag	360
agaagatato ag	accadata ==	yuctocaaa t	ccagaaaac	aaggactatg	aagagccaaa	420
gaaagtacgg aa aactgaaaat ti	taadctaa e	couggacat t	ttccttgct	ttttgattta	tttaggggac	480
aactgaaaat tt acccacaaat aa	atatatora ~	yaaradaga g tataattta -	yctgaagaa	gactggcttc	actgattatt	540
acccacaaat aa gttattcact ta	aacagattt o	egragiliy g	ayayaatti cataacete	clagattta	atataccaaa	600
	ge	caayee a	cataayetg	aattecatag	acgagattgc	660

atttatettt ttattaatat eetgaegttt etttttggaa teaagtatgt aaataeettt	
	720
The state of the s	780
	840
The state of the s	900 960
The state of the s	1020
	1020
TO THE TOTAL CALL BACKBOLD AND A SOCKAST AND ASSOCIATION OF THE PARTY	1140
	1200
The state of the contract and a cont	1260
Jameselle auditilates arantettas assassas .	1320
	1380
catgicatti agcitatti caattigata gillaatggci ataaaaaatt ttattigtaa	1440
atatatggta ccataaacca aaacgtttat gttttgctgg gagatcattt tagatgtatc tttgtgccta tctgtgaaaa tttccttaga attcttaaag taaatttgct gaattggatg	1500
taaaagttta aaagacagta gatacacatt gttaaattgt tcttcagaaa agtgattctg	1560
tttacttcct ctgaaatac ctctttccct gtacctcacc caagtgtgta ttaccatttt	1620
	1680
STOCKER CLLLADADAC AFRRECECSO FREEERS	1740
	1800
J J J J J J J J J J J J J J J J J J J	1860
	1920
"J'''J''' LLCALCEER OF BOOK SAN ANALEL	1980
	2040
	2100 2160
	2220
	2280
	2340
	2400
	2460
ggtcacagaa tgaatttcag tcattgctta ggtaagcctg gtaaaaaaaa aaaaaaacgg gattccatga ttagtgtgtt gaaaaggggg ttgccaaagt gtaacagaca gagccgggtg	2520
	2580
	2640
	2700
	2760
ctccagcctg gccacagagt gagattccat ctcaaaaaaa aaaaaaaaa aaaaaa	2820
· · · · · · · · · · · · · · · · · · ·	2876
<210> 23	
<211> 1052	
<212> DNA	
<213> Homo sapiens	
Supremo Supremo	
<400> 23	
togacccacg cytocycca cycytocycc cacycytocy gycygycyca gyacytycac tatyyctcyg gyctyrath godach garranada gydynyddiae	
	60
	120
	180
	240
	300
	360
	420
	480
	540 500
	600 660
	720
	780
ggggtcaccc tggggggtta gggacctatt tttaacacta gggggctggc ccactaggag	840

WO 00/06698 PCT/US99/17130

ggggagagt attttaaa	t tggagggag	g ggtgggggg a aaaaaaaaa	g gggaaaaaaa a aaaaaaaaaa	a aataaaaaa	a tatttattt a aaaatttta a aaaaaaaaaa	900 960 1020 1052
<210> 24 <211> 154 <212> DNA <213> Home						
<400> 24						
	t tcctcgatcc	tactctgtct	ctctttgttt	tggagaagtt	caatttacca	60
gctggatat	g tgggactagt	attcctgggt	atggcactgt	cctatgccat	ctcttcacca	120
ctatttggt	c tcctaagtga	taaaaggcca	cctctaagga	aatggcttct	ggtgtttggc	180
aacttaatca	a cagccgggtg	ctacatgcto	ttagggcctg	tcccaatctt	gcatattaaa	240
agtcagctct	ggctgctggt	gctgatatta	ı gttgtaagtg	gcctctctgc	tggaatgagt	300
ttaagtaga	taaaaatta	aattctcagt	tgtgcacatg	aaaatgggtt	tgaagaggga	360
atgggacca	cgctgggtgg	atttctgtat	. cctagtgcaa	tgtggtcaat	tggtgctttt	420
atacaaggt	tatgggctct	gataagtgga	. gagaaaaccg . ttagccatgg	gcttgtttta	tctactagec	480 540
tattcaagga	gaaaaaggtc	taaatctcaa	aacatcctca	gcacagagga	ggaacgaact	600
actctcttgc	ctaatgaaac	ctagtccgat	ggatcctgga	ttgatacaag	gttgagaaat	660
gaatgctcct	ggccttaaac	atcaccgtag	gaagggtttt	taaaatttta	cgcgcaaaac	720
tccgtggacc	ccgtgccagt	gtcttggaag	tgtcaacgtg	tttttggatg	atcctgtatt	780
gggctgtact	tactgtgata	ctgaaaagct	gtcctgctga	agcagctata	tttgaaatat	840
CCasacttat	aggagtaatt	aaaaacaagc	aaaacaaaac	aagacttagt	ttttaaatga	900
ttttttatto	ccttaaagat	gttgttatta	actcgagtta	gttcttattt	cctctgttta	960
atgaattaat	taagtacact ttgatatgct	ctaaatacat	atgracett	canatatat	agggaaagaa	1020
ctatgaaatc	agtttttaaa	agatatactt	tetettete	ctanagatttt	teggtettet	1080
tcaaaaggaa	gaattettge	ctgccataca	gaaactctct	aggagtcct	gaccttaage	1140 1200
ttttctaaaa	attctgtttg	tgtgaaaagt	acaagaataa	caatacttac	aacttccatt	1260
tttgtaacct	acgttcactt	atgatctgga	tttataaaca	ttacttggta	taacgttttt	1320
catttccttt	aatgtctctg	ttttttggct	ctaccatctg	ttttgtttt	gtttttatct	1380
atatcttggt	agatgtattt	catccctaga	gcaggtcagc	ctccttcccc	taatgcgaat	1440
gcttgttttg	ttagggaagg	gcttcctcca	acttcgtgtg	aaattgtgat	gttgaagtga	1500
ataaatgtct	attgtgtaac	aaaaaaaaa	aaaaaaaaa	a		1541
<210> 25						
<211> 2079						
<212> DNA						
<213> Homo	sapiens					
<400> 25						
ggcacgaggg	aggcgcggct	gcgggacctg	actagattct	acgacaaggt	actttctttd	60
catgaggatt	caacaacccc	tgtggctaac	cctctgcttg	catttactct	catcaaacgc	120
ctgcagtctg	actggaggaa	tgtggtacat	agtctggagg	ccagtgagaa	catccgagct	180
ctgaaggatg	gctatgagaa	ggtggagcaa	gaccttccag	cctttgagga	ccttgaggga	240
gcagcaaggg	ccctgatgcg	gctgcaggac	gtgtacatgc	tcaatgtgaa	aggcctggcc	300
cgaggtgtct	ttcagagagt	cactggctct	gccatcactg	acctgtacag	ccccaaacgg	360
dattattoto	tcacagggga	tgactgcttc	caagttggca	aggtggccta	tgacatgggg	420
gartattacc	atgccattcc	acggctggag	gaggctgtca	gtctcttccg	aggatettae	480
gcttatttcc	agacagagga	tatttatt	ccagaagatg	totata	cttggccttt	540
tacagecean	gggcaggaaa ataataagag	astaacceaa	aatatattaa	aatatossa	gittetete	600
gagagcccca	accacgtggt	agctgaggct	atcatccage	aacacyaaag	acccaaceta	660 720
cagaccagag	acacctacga	ggggctatgt	cagaccctgg	gttcccaacc	cactototac	720 780
	÷		5	J =		700

```
cagateceta geetetaetg tteetatgag accaatteea aegeetaeet getgeteeag
                                                                       840
  cccatccgga aggaggtcat ccacctggag ccctacattg ctctctacca tgacttcgtc
                                                                       900
  agtgactcag aggctcagaa aattagagaa cttgcagaac catggctaca gaggtcagtg
                                                                       960
  gtggcatcag gggagaagca gttacaagtg gagtaccgca tcagcaaaag tgcctggctg
                                                                      1020
 aaggacactg ttgacctaaa actggtgacc ctcaaccacc gcattgctgc cctcacaggc
                                                                      1080
 cttgatgtcc ggcctcccta tgcagagtat ctgcaggtgg tgaactatgg catcggagga
                                                                      1140
 cactatgage etcactttga ecatgetacg teaceaagea geceeteta cagaatgaag
                                                                      1200
 tcaggaaacc gagttgcaac atttatgatc tatctgagct cggtggaagc tggaggagcc
                                                                      1260
 acagcettca tetatgecaa ceteagegtg cetgtggtta ggaatgeage actgttttgg
                                                                      1320
 1380
 ctggtgggag ataagtgggt ggccaacaag tggatacatg agtatggaca ggaattccgc
                                                                      1440
 agaccetgea getecageee tgaagaetga aetgttggea gagagaaget ggtggagtee
                                                                     1500
 1560
 gcctcctgga agaaggcctt gtcagctttg tctgtgcctc gcaaatcaga ggcaagggag
                                                                     1620
 aggttgttac caggggacac tgagaatgta catttgatct gccccagcca cggaagtcag
                                                                     1680
 agtaggatgc acagtacaaa ggaggggga gtggaggcct gagagggaag tttctggagt
                                                                     1740
 tcagatactc tctgttggga acaggacatc tcaacagtct caggttcgat cagtgggtct
                                                                     1800
 tttggcactt tgaaccttga ccacagggac caagaagtgg caatgaggac acctgcagga
                                                                     1860
 ggggctagcc tgactcccag aactttaaga ctttctcccc actgccttct gctgcagccc
                                                                     1920
 aagcagggag tgtcccctc ccagaagcat atcccagatg agtggtacat tatataagga
                                                                     1980
 ttttttttaa gttgaaaaca actttctttt ctttttgtat gatggttttt taacacagtc
                                                                     2040
 attaaaaatg tttataaatc aaaaaaaaaa aaaaaaaaa
                                                                     2079
 <210> 26
 <211> 1947
 <212> DNA
 <213> Homo sapiens
<400> 26
tgtaaacaga ttggagaatc tagcaataag attcaaagct aatctggagc ataaaggcac
                                                                      60
agttcagaga cagaataaca gggatcacaa gcatgaatta aaaggaattt atttgcttca
                                                                      120
agttcctaga tacaaccttc ccatgctgca cttctccact gtcggagcac gttccgaaaa
                                                                      180
acagaatgcc ttgatccctg gtgggtgcga aggcagttgt tagggatggc aggcattggt
                                                                      240
gggctccaaa agatgaaggc cccacacaca ggtgtgctgc atttgggatc tgtgtgggtg
                                                                     300
tttettggae cetttettet gggagtaggg tacacactaa egtttaatee getgtetggg
                                                                     360
tgcatgtcca cagtacggtg gctaaactcg aacatcactg caaataggac gctgagcagg
                                                                     420
teegtetgte atgteacgee actgeacagg teettgteec cacaegaegg ggagtaettg
                                                                     480
cgtcagatgt tattgaatag ctcgtctcgg gcaggggaag cggggagttg gggatattaa
                                                                     540
ttgggggttt taattctatt atcatgtcag ctgacattat gactatataa tgtagttaga
                                                                     600
gacaattttt atcttgctta tagtaaaggt tcagcctgcc aattgtaaat cattctaatt
                                                                     660
tggcaggctt atttttgaca ttggaaaggg cagaaagcga tttgccccag tagtgtaata
                                                                     720
ggagttatag accagagget gaaacccaaa ctatataaaa aggaattcag tggagggggc
                                                                     780
tttgtaatct ccattaattt gtgttgctac ttccaggatc accaaaaatt acatgtaatt
                                                                     840
ttacatgtta aacacattga aacataacct atgtttataa agcataacgg gcttcccttc
                                                                     900
cagaagetet cetgettgte atgaagtgag aacaatgaaa agteatagea gataeteagt
                                                                     960
ttaactctgt gtagaaccta gtagtgtttg agctgttatt cagatttgaa ttcagactgt
                                                                    1020
gtgttgtttg cttatggaca ctgcctgtcg ttctgtcact gttaaattaa tgagtctata
                                                                    1080
aggtttttct tccagaggcc ataggtgaca tcactaaaat tgcaagataa attgtaatct
                                                                    1140
ttgctgctgc tgcactcccc aacctctccc ccaccccccg tggtgtgctg ctttctagat
                                                                    1200
gagcgtgttt tggagcaggc ccatctggga cactctatgc tttcaccaag gaagtgcgat
                                                                    1260
ctgagcagcc acaatccagc caaaagagga tcgtagatat ttgctctgat caactagatg
                                                                    1320
aaaatatagc agaatggatt tagcccactg ctctgtttta tccaactgag tctctgacca
                                                                    1380
gcaattggtg cataattatt acagcaaaag ttaagaaatg aaactgtagc aattatgtaa
                                                                    1440
atgaatgtgt tggcctctta atacctgtta ctagtggact tcctgtgagg aagttagttt
                                                                    1500
tttgttttga tgaaatgctt tcgtttttta aatcttaatt ctgctgtcca catcctccca
```

aagtgtgctt acttcatttg tttaatttaa atgaactttc ctccttgtat gtatgaggtg

aggeeteaaa ggaeetttee tttaggteat atteeteaga aagtetteaa tetteeettg

: -<u>:</u>-:

1560

1620

1680

<pre><210> 27 <211> 3379 <212> DNA <213> Homo sapiens </pre> <pre><400> 27 atgatattca agtegaattc ccctcactaa agggaaaaaa getggagetc cacegoggtg geggegecetc taaacetact ggatececeg ggatgcagga atteggeacg aggtetegge 120 tecetecegge accegecegg cecgetyet cccggetect eccggeateg gggagetyeg 120 geagegeget getgetetyg ggetgaagg gggaggaggaaggaggaaggagaagaggagaaaactgt gggagaacca cacegogggtg agaactgag gggagtaggtg ggagaagatggggaaaactgtg ggaaaactgt gggaaactgg gggagaacaca cacegoggtg agaactgaat gggaatgagtg ggaaaacaca ggaaaccac accegoggtg agaatgactg gggaaattg gggagaacaca cacegoggtg agaatgaac gggaaactga tgtgetetaca 360 gtagaaacac ggaatctgact tgtegaaga cacaaagtga acceteggat gagaactgat gtgetetaca 360 gtagaaacac ggaatctgact tgtegaaga aaaagatcac gatagactgat tgtetetaca 360 gtagaaacac ggaatctgact tgtegaaga agaceacac cacaagtga gagactgact tgtetetaca 360 gtagaacac ggagattgat tgtegaagaga agactaca ctteeggtt ectecacactgat 420 etcaccttgat tagectggaa gatgatgtt caagaaaga gagtatgatg gtgaagatga gggaagacgat gggagagat tgtegaagaga agactataa gtaceccaa gagacttat tggactgaa gagaagaaga ggttattt caacactgat tgtecacacaa gaagaagaaga agattattt caacacttit tgcctgaa gaagaagaaga agattattt cacacactat tggcatgaa gagaagaaga agattattt caacacttit tgccacacaat ttacacacata gaagacaaga agattattt caagagtag gagaagaagaagaagaagaagaagaagaagaagaaga</pre>	aggactatti agatatatai	c agaagtata	g gctgtcgttg a agatggtgt	g gcggcagcag	g tatattctga	a aatgtctcat a aatgtctcat a aaaaaaaaac	1800 1860 1920 1947
atgatattca agtegaattc ccctcactaa agggaaaaaa gctggagctc caccgcggtg gcggccgctc taaacctact ggatcccccg gatgcagga attcggacg aggtctcggc 120 tccctccggc acccgcccgg ccggctgcc cccggcccct cccggccat ggagactgcg 180 cgcggctgtc gctgctctgg ggctgcacgg tggtgccgc aggactgatg ggagatagtg 240 gagaaaactctg ggcagacac acctgcggtc agaatgctac cgaactgat gcgaattgg 240 gaaaactctg ggcagacac acctgcggtc agaatgctac cgaactgat tgctgctatc dgtgagaacac ggatctgct tgtcggcag caaaatgtac cgaactgat tgctgctatc 420 ctcacctgg tcacctgca ttgtcgcatg agaatgcaa catctgcggt cacctgcat tgtcgcatg agaatgcaa ctctcagctg tcacctgca ttgtcgcatg aaaagccca ggagatagt tgtgacaag aaaagctca gttagactg gagatgaat 540 ctcacctagt tgcggagaa gtgacaaga aaaagatca gttagactg gaagtgtgat tctacctact tgtgaggaa gaccttaaa agcccccaa ggacttata tgcgaggaa gaccttaaa agcccccaa ggacttattg acttctagac 660 accgctacaat tggcctggaa gatgatgttg tcaagaaag gccggctgc atggtgctgg 600 accgctccaa ggacttaac agtgcaaaag agcttatat agtactgcaaga gaccggctc tttttcatga atggcaaag tcaggagaa gcgcggctc ttttacatga atggcaaag tcaggagaa gccggctgc tgcttataa agtgcaaaag tcaggagaa gcgcagactg tgtgaaaga acaccttaa agtgccaaag tcaggacaaga gcgcagactg tgcagaaga cacacccac aggactgac cacacctgc gaactgactg tgcagaaga cacactacta tgaccaatg cacacctgt catggctaca agaactgac gaacactgc gaacactgc gaacactga gacacttaa caataccaga tgattatat tacacacta tgattcatgg aatgggaag acacttaca acattccaat tgattcatgg aatgggaag acacttaca tggccaaag acacactga gacactgac gaacactgg gaacactgg cacacagagag cacacacag agtgaaac ctgcaagtg aatgggaag cacacacag agtgaaac ctgcaagtga aatgggaag cagacactg ggaacactg ggaaacactg ggaaactga ctgcaaacac gaacaccag agtgaac ctgcaagtga aatgggaa ctgaacacga gacacacag agcacacta ggaaacac ttgcaaacc gtgaaacg ctgcaacaca acaagaag gacagac ttgcaaacac gtgaaacac gaacacacaca aacaacagag gacacatata cacaacaagag gacacacacaaga gacacacaca acacacaga gacacacac	<211> 3379 <212> DNA						
Geggecget elaaacetact ggateceeeg ggatgeagga atteggeag aggeteegg 120 cecteegge accegeeegg eeggeteeg gegegeteg teggtgeee elageteegg 180 cegegetet getgeteeg gegeteeg teggtgeee aggactagt ggagtagetg 240 gagtgagtte cegetgtgaa aaageetgea acceteggat gggaaatttg gettteggge agaaacettg ggeagaace acetegegte agaatetae eggaetgtae tegetetaea 360 gtgagaaca ggatetgaet tegegeage caaatgtga caagtgeaat getgeetate ctcacetgge teacetgee tetgecatgg cagacteate etteggat ggagtagat ggtggagte tegegaggat gtgeacagag aaaagateea gtaagtgeat getgeetate ctcacetgge teacetaet gtgeatgte agageetaea gttaageeg agagetgeet ggtggagte tegegaggat gtgeacagag aaaagateea gtaageetg agagetgaat 540 ctcacettee teacetaatt gtgatgttea agteeceea geeggetee atggtgetgg 600 accgeteeca ggaetttegga aaaacatgga agcettataa gtaattteg actaceatge 660 cegetacatt teggeetgga gatgatgttg teaagaaagg egetatttt catteecaa tegeeteetae accgaaagag aggetattt caagagaag egetgagate accgaaceatg accgateecaa gaateettae getacetaat gatteetgaga gaatgageeg egetgeed acggatgege egetgeed accgaacatt teacacata tegecateta gatteeatg cacagagaga getgaagate accaaceate acctaacaat tegecateta gatteeatg cactggaagg egetgaagat accaaceate acagaaagga ctgacaagat catacetgt catggetea gacetgteaa gacetgetaa acattecaac teggecateg gaagtgtatg tetteggaaca acacagaaga agcecatge aatggecaeg ctgateaatg catacetgt catggagatg egegeagga egecatge acattecaaca teggecateg gaagtgtatg tettegaga acacagaaga agcecatge acattecaaca teggeacate deacetgt catggagata cacacagaaga agcecatge acattecaa teggeacate deacetgt atggagetae gacetgteaa gacecatge acacacagaag gacagate egeagaggt atggatggat ettacetga cacacagaaga acacagaaga gaceactge gtaacatg geagaac ettagaaceg categgagat ettacetga cacacagaaga cacacagaaga gacacatge gtaacacta gacacagaag gagaagaga ettacacaca tetacagaa acacaagaaga gacacacaca ettacacacacacacacacacacacacacacacacacaca	<400> 27						
ccctccaggc accegacy ccggactgct cccggactect cccggacatg ggaagactggg gaggtagattc ccgctgtgaa aaagactgaa acctcagat ggaagattgg 240 gagatgagttc ccgctgtgaa aaagactgaa acctcaggat ggaaatttg gctttggggc gaaaaacttg ggaagacac acctgggat gaaatgacta cggaactgtac tgcttctaca 360 gtagagaaca ggaatctgact tgtcggatga caaaatgtga caaagtgaat gcttgctatc 420 ctcacctggc tacactgact tgtcgcatgg caaaatgtga caaagacatg gaagatgactg ggtggaaac ggactgact tgcgcatgg caaaatgtga caaagacacg gaagatgactg tcacctggc tacactgacat gtgcacagag gaagactgac ctctcggacat 480 ggtggaagt tgcggaggat ggtgacagag aaaagatca gttagacctg gaagctgaat tctacttcac tcacctaatt gtgatgtta agtccccag gccggtccca ggactttggg aaaacatgga agccttataa ccgctaccat tggactggaa gatgatgttg taaagaaaga ggctatttt caccaccat 1720 actccaqtcc ttttccatga agtgcaagag aggttattt caaagaagag cgctatttg acttctaaat 720 actccaqtcat tttacaacta agtgccaaag ttcaggaca gcgaagata cacaacctcc gcgtgaaca gaacccttaa agtgccaaag ttcaggaca gcgaagata cacaacctcc gcgtgtgage tgctgaaaag aagtgttgt ccttgtcaga gaaatgact gaacgaagag 900 cctcaacatt ttacaacta tgccatctat gatttcattg tcaaggaca gcgacagaga gacgcaatgag aatggccacg ctgatcaatg catacctgt cattggataa gaccgtaatg gaacgaagag 900 cctcaacatt ttacaacta tgccatctat gatttcattg tcaaggaca ggaccatgg aatggcaca ctggacaag gaaggtatg tgttagacaa accacagagg cagccattgc acattccaca tggtccatgg gaagtgtatg tgttagacaa accacagagg cagccattgc acattccaca tggtccatgg gaagtgtat gttgtagacaa accaagaag cagccactgc gctccaacg ggaagaac ctgcaagtgt agtggagaga gactgtatg gaaaaagaggg gctcccaaaga gacgagaac ctgcaagtga aggaacagga ctgataatg gaaaaagagg gcttccaaaga gacgatata caaagagaga aggatgat tctgtgatga ctgtaacagag gcacactga ggaagaat gaggaatag aggagatg aggagatga ctgataacagag gacacatga gagaacagag ctgtaacagag gcacacaga gacgatata caaagagaga aggatgata tctatggaga gacacatga ctgcaaacag ccttctcaa ctcagatga ggagaacag aggagatga caacaagag gacacagag ggagactga ggagactg gagagatga gacagagaga gacagaga ctgaacacagagagagagagagagagagagagagagagag	atgatattca	agtcgaattc	ccctcactaa	agggaaaaaa	gctggagctc	caccgcggtg	60
ggggggtgtt gctgctctgg ggctgcacgg tggtggccg aggactgat gggaatagtt gagtagctg gagtagatt gggaaactt gggaaactt gggaaactt gggaaactt gggaaactt gggaaactt gggaaactt gggaaactt tgtcggcag caacttgaat cgaattgtat cgtctctaca 360 gtgagaacac ggatctgact tgtcgcatga caacttgtac cgaactgtac tgttctcaca 420 gtgagaactt tgtcgcatga caastagtaa caastgtaa gctgcttat 420 gtgggagat tgtgcacaga aaaagatcca gttagacctg aaggtgaat 540 tctacttcac tcacctaatt gtgatgttca agtccccacag gctgctgca atggtgctgg 6600 ccgctactt tggcctgga aaaacatgga agctttaaa gtactttgg actaactgct 660 ccgctacatt tggcctgga gatgattgt tcaagaagaag cgctatttg actccaactg ccccacag gaccttaca agtgccaaag ttcagaagaa ggctgatatt tcaacacta ttcaccacta tgccacacta tcacgaagta actccaagta actggaagaa gagcttatt caaagattga accaactcc 840 cgcgtgcag tgctgaaaag acgtctttgt ccctgtaaa gaaatagccag cctgaaagaag gaccttaaa gtgccaacag ttcagaagaa agttctgt caaggacaa gccgaagaaga 900 cccaacact ttaacacacta tgccacacta catgccacta atgtccaagt daaacacatt taacacacta tgccacacta caatgccaga gaagttatt gttaagaca acacacagaag cagccactgc 1080 aatggcaag cccgtatata aatgccag cccgttaata caatgacag gaagtgatat gttaagaca acacacagaag cagccactga 1020 aacacactgt cccgttata caatgacgg caatgagaaga gacagtaatg aatgagagag cagcactga gaccactga acacacagaag cagccactga 1080 gctaaatgtg ggaaggacat accaacagaag cagcacactga 1080 aacacagaaga gacagtatt ccagaaggt aatgggggggggg	geggeegete	taaacctact	ggatcccccg	, ggatgcagga	attcggcacg	aggtctcggc	120
gagtgagttc cecetytgaa aagecetga acceteggat ggaaatttg getttgggge 3a6 gaaaactet ggeagacac acctegggte agaatgtaa cgaactgtaa tgettetaca 360 gtgagaacac ggatetgaet tgteggeage caaaatgtga caagtgaat getgectate 420 ctcacetgge teacetgeca tetgecatgg cagacteate ctteeggtt cetegaact 480 gtggagagte teacetgagat tgeegaggat geaaaggatea gtagaactgaat 540 cccactact tgeetggagat gtgacacgag aaaagateaca gttagaactg gaagtgaat 540 cccgctacat tgeetggaa aaaaatagga ageettataa gtaetttegg actaactget 660 accgeteca ggaetttggg aaaaactgga ageettataa gtaetttegg actaactget 660 accgetecat tgeetggaa gatgatgtte teaagaagga gegetatttg actaettgga accactact cgeegtgaact tteeagaagga aggetattt caaagettg teaceacat 780 accacate tttlecatga actgaagga aggetattt caaagettg teaceacat 780 acgatacaga gaaceettaa aggecaaag teagaggaa gegetattg teacagaagag 900 cctcaacatt ttacacacta tgecatetta gatteattg gacetgaagat accaacetee 840 acgagcacg ctgateaatg cataectgt catggetea gacetgaagga cegagagaag 900 cctcaacatt ttacacacta tgecaatgt catggetea gacetgaagag cagacatgge acattecaca tggtccatagg gaagtgtatg tgttaagaagg cagactgaagag gaceceaagga 1020 acattecaca tggtccatagg aaatggaac catgggagga catggagagga cagacatggaac cagaggaac cagaggaga cagaggaga cagaggaga cagaggaga cagaggaga cagaggaga cagaggagaga cagaggagagaga	teceteegge	acccgcccgg	cccggctgct	cccggctcct	cccggccatg	gggagctgcg	180
gaaaactctg ggcagacac acctgcggtc agaatgtac cgaactgtac tgcttctaca gtgsgagacac ggaactgact tteggacag ccaaatgtga caagtgcaat gctgcctatc ctcacctggc tcacctgact tctgccatgg cagactcatc cttccggttt cctcgcacat 480 ggtggcagtc tgcgagagga gtgsgatga ggcacagag aaaagatcca gttagacctg gaagctgaat tctacttcact cacctaatt gtsgtyttca agtccccaad gccggctgcc atggtgetgg 600 accgctccca ggactttggg aaaacatgga agccttataa gtactttgg actactggt 660 accgctccca ggactttggg aaaacatgga agcctattaa gtactttgt actcctaact 720 accccatact ttggcctgga gatgatgtt ccaagaagg gccgtatttgt accacacat 780 accccaqtcc ttttccatgc actggaagga aggttattt caagaagga gcctatttgt accacacat 780 accgcgtgcag tggtgaagga acgtcttat 2 gattccagag gaaatgacct gacgagagg 900 cctcaacatt ttacacacta tgccatctat gatttcattg tcaagaggac gccgaaggag 900 cctcaacatt ttacacacta tgccatctat gatttcattg tcaagaggac ctgcttctgc 960 aatggcacag ctgstaaacg acagtcatg catggcagag cctgttcaagag cctgcaagga 1020 aatggcacg ctgataacg acagtgtatg tgttagacca acacacagag agcccaagag 1020 aacacacaca aggtcaagaa ctgcaagtgt atgtgggggg cagctgtg gacacaggg 1140 gctcccaaca gatgcagaac ctgcaagtgt atgtggtggtg cagctgtgggagacg accgctgtg gacagagagacg agcagtattg ccagagggagacg agcgtgtggggaggacg agcagtgtg gcaagagagagacg acccccccccc	cgcggctgct	gctgctctgg	ggctgcacgg	tggtggccgc	aggactgagt	ggagtagctg	240
dtgagaacac ggatetgact tgteggeage ceaaatgtga caagtgeaat getgeetate 480 ggtgycagte tgeggaggat gtgeacaagag aaaagateca gtteggetgad 540 tetaettecae teaectaatt gtgatgttea agteeceag geeggetgee atggtgetgg 600 aeeggeteeca ggaetttgg aaaacatgga agcettataa gtaetttgg actaetget 660 cegetacatt tggeetgga aaaacatgga agcettataa gtaetttgg actaetget 660 cegetacatt tggeetgga gatgatgtt teaagaaagg egetatttt caaagcatg acteatactge actggaagga aggtatattt caaagcatga geetataag accetae aggeetaag acgetaagga gatgatgatg teaagaagga geetattt caaagcatg taetggaagga gegggaga geggagaga geggaagga geggaagga geggaagga geggaaggag gegaaggag geggaaggag geggaaggag geggaaggag geggaaggag gaggaagga	gagtgagttc	ccgctgtgaa	aaagcctgca	acceteggat	gggaaatttg	gctttggggc	
ggtggcagtc tggcgaggat gtgcacagga cagactcatc cttccggttt cctcgcacat 540 tctacttcac tcacctaatt gtgatgttca agtccccag gccgctgcc atggtgctgg 600 accgctccca ggactttggg aaaacattgga agtccttataa gtactttgg actacatggt 660 accgctcacatt tggctggaa gatgatgttg tcaagaaagg cgctatttgt acttcctaaat 720 actccagtcc ttttccatgc actggaagga aggttatttt caagctttgt actccaacat 780 acgatacaga gaacccttac agtgccaaag ttcaggaga ggtgaagacc gacacatcc 780 acgatacaga gaacccttac agtgccaaag ttcaggaga ggtgaagacc gacacactcc 6840 cgcgtgcagc tgctgaaacg acagtcttgt ccctgtcaga gaaatgacct gaacgagag 900 cctcaacatt ttacacacta tgccaatcat gatttcattg tcaaggagac gcccagga 1020 cctcaacatt ttacacacta tgccaatcat gatttcattg tgatagcac acaacagaga gacccatgc 1080 cagcactgg cccgttata caatgaccgg ccatgggagg cagctgatgg cagctgatgtg gaaggtgatg gttagaaca acaagcagg cagccactgc 1080 cagcactgtg ccccgttata caatgaccgg ccatgggagg cagctgatgg cagctgatgg ggaggagaccg aggaggagaccg aggaggagaccg aggaggagac ctgcaaggg aatgggggggggg	gadaactctg	ggcagacacc	acctgcggtc	agaatgctac	cgaactgtac	tgcttctaca	
ggtggcagtc tgcgaaggat gtgcacaaga aaaagatcca gttagacctg gaagctgaat 540 ccactctcac tcacctaatt gtgatgttca agtccccaag gccggctgcc atggtgctgg 600 accgctccca ggactttgg aaaacatgga agccttataa gtactttgcg actaactgct 660 ccgctacatt tggcctggaa gatgatgttg tcaagaagg cgctatttgt actactccagtc tttcatcatg agtgccaaga gagttattt caaaagcttgt tcacacacat 720 accggtgcagc tgctgaaacg acagtcttgt ccctgtcaga gaaatgacct gaacgaagag 900 cctcaacatt tacacacta tgccatctat gatttcattg tcaagggcag ctgcttctgc 960 aatggccacg ctgatcaatg catacctgtt catggcttca gacctgtcaa ggcccatgga 2 cacattccaca tggtccatgg gaagtgtatg tgttagcaca acacagaagg cagccattgd cagcacttgt ccctgaagaga catggaggag cagctgtgg cagcaaggag 1020 acattccaca tggtccatgg gaagtgtatg tgttagcaca acacagaagg cagccatgg cagcactgg caggaagtcg catggaggag cagtgatgg cagcactgg cagcactgg cagcactgg caggaagtcg catggaggag cagtgatgg cagcactgg cagcactgg cagcactgg cagcactgg cagcactgg cagcacatgg caccagaagag gaagtgatg catggagag cagtgatgg cagcactgg cagcactgg ggaacacaaga gacaattatg cagaggatcg agtgctgdg tctgtagaag caccacaaga gacacacaaga gacaattag cagaggatcg tgtcacagg ttctgtagaa cttcgtgac cttcctgcac cacagagtgc ttctcctgca ctcctcctcaag ctccagatgg ttgcaaaacca gatgacgt tgcaagagga ttctactgga cttgcaagaga 120 ccttcctgca actcagtgac cttctgcgac catggaggat tctactgga cttgcaagaga 122 ccttcctgac actcagtgac cttctgcgac catggaggat tctactgga cttgcaagaga 122 ccttcctgac actcagtgac cttcaggaga tgtgacgtac ttgcaagacct 1440 ggctgtgcag gcgaggatgt tgcaagagga tgtgacctac tccaggaga actaggagatcat 1500 ggctgtgcaa gcgaaccag gagaatcga gatgaagaa actggagaacacaa gacacacaa gacacacaa gacacacaa acatagactg gaagaacacaca aaagacacaca gacacacaca acacacac	gragadadad	ggatetgaet	tgtcggcagc	ccaaatgtga	caagtgcaat	gctgcctatc	
accgctccae gacetttagg aaaacatgga agccttataa gtactttgg actacatggc 660 ccgctacatt tggcctggaa gatgatgttg tcaagaaagg cgctatttgt acttctaaat 720 actccagtcc actccagtcc ttttccatgc actggaagga aggttatttt caaaggtta tcaagaagga 900 cctcaacatt ttacacacta dgccatcatt gatttcattg tcacagagga 900 cctcaacatt ttacacacta tgccatcatt gatttcattg tcaagaggaa ctgctttagc actggcacg ctgctaaacg acagtcttgt ccctgtcaag gaaatgacct gaacgaagag 900 cctcaacatt ttacacacta tgccatcatt gatttcattg tcaagggcag ctgcttctgc acattccaca tggtccatag gaagtgatagt tgttagcaca acacaccagga 1020 acattccaca tggtccatag gaagtgatagt tgttagcaca acacagcagg cagccactgc 1080 cagcactgtg gcccgttata caatgaccgg ccatggagg cagcactgtg 2020 cctcaacatt gaggagac ccagcaagg 2020 cctcaacatt gaggagac ccaatgagag 2020 ccccaacag aggaggaca caatagccg 2020 ccccaacag aggaggaca caatgaggag 2020 ccccaacag aggaggaca 2020 ccctcaacagag 2020 cccctacacag 2020 cccctcaacagag 2020 ccccaacagag 2020 ccccaacag 2020 ccccaacag 2020 ccccaacag 2020 ccccaacagag 2020 ccccaacagagagacag 2020 ccccaacagagagacag 2020 ccccaacagagagacagagagagagagagagagagagag	gatagaaga	taaaaaaaa	cetgeeatgg	cagactcatc	cttccggttt	cctcgcacat	
acceptance aggaettegg aaaacatgga agcettataa gtacttegg actaactget 720 actocagted tegeorgga gatgatgtt toaagaaagg cgctatttgt cacacacat 720 actocagted tegeorgaa gatgatgttg toaagaaagg cgctatttgt cacacacat 720 actocagted gaaccettac agtgocaaag toaagtgttg toaagagaa geggaacac 320 acgettacacat toaacacat cacacacat cacacacat toaacacat toaacacat toaacacat toaacacacat cacacacat cacacacat toaacacacat cacacacacat gaacgaagag 900 acattcacacat togocatcat gattcattg caaggagac gacctgtcaggag cagccacacacacacacacacacacacacacac	tctacttcac	tcacctaatt	graceagag	aaaagatcca	gttagacctg	gaagctgaat	
ccgctacatt tggcctggaa gatgatgttg tcaagaaagg cgctatttg acttctaaaat 7200 acgtaccagtc ttttccatgc actggaagga aggttattt caaagctttg tcaccaccat 7800 acgatacaga gaacccttac agtgccaaag ttcaggagca gctgaagact gacagaagag 9000 cctcaacatt ttacacacta tgccaacat ttacacacta tgccatcat gatttcattg tcacaggagca ctgctctcg 9600 aatggccacg ctgatcaatg catacctgtt catggcttca gacctgtca gaccaggag 1020 acattccaca tggtccatgg gaagtgtatg tgttagcaca acacagcagg cagccactgc 1080 cagaactgg ccccgatga gaagtgtatg tgttagcaca acacagcagg cagccactgc 1080 cagaactgtg gggaggact cagaggagac cccatgggaagg cagtgatggt aatggggag cagtgatgg cagaacgggg cagcactgc 1200 gttaatggtg ggaggactc agggaaccg ctgcaaggg agggaatcg agggaatcg agggaatcg ctgcaaggg cagcactgg caaacagagag 1200 gttaatggg gacggaac ctgcaagggg aagggaatcg ctgcaagggg cagtgatgggg cagcaggggg cagcactgg 1200 gttaatggg gacggaacggg cagcagggg cagcaggggg 1200 gttaatggg gacggagac ctgcaagggg aagggaacg ctgcaaggggg aagggagggg aagggaggg aagggaggg	accontroca	ggactttggg	grgargrada	agreecedag	geeggetgee	atggtgctgg	
actoccagtcc tittecatic actigaagga aggitatitt caaaggitit cacacacta 3840 cggatacaga gaaccottac agtigcaaag titcaggagga gaatigactit gaacgagaga 900 cctcaacatt tidacacata tigccaatcat gattitatit caaggigaa cacaggagaga cigcitititie gaatiggacaagt titacacacta tigccaatcat gaatititie tidacacata tigccaatcat gaatititie tidacacata tidacacata tigccaatcat gaatititie tidacacata tigccaatcatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatatacacatatacacatatacacatatacacatatacacatatacacatatacacatatacacatatacacatatacacatacacatatacacatatacacatatacacatatacacatatacacatatacacatacacatacacagaagacacacac	ccactacatt	tagectage	aaaacacyya	taaaaaaaa	gracitigeg	actaactgct	
acgatacagagaaccettacagtgccaaagticaggagcagetgaagatcaccaacetce840cgggtgcagctgctgaaaagacagtettgtccctgtcagagaaatgacctgaacagagag900catacacatatttacacacatatgctacatetgattcatetgattcatetgacctgtcaaggcccagga1020acattccacatggtccatgggaagtgtatgtgttagcacaacacagagagcagcccatgc1080cagcactgtgccccgtttatacaatgaccggccatggaggcagacctgtgcaacacgggg1140gctccaacagagtgcagaacctgcaagtgtaatgggcatgctgtaacctgtacttcgac1200gttaatgtgtgggaggcatcagggaatcgagtgtgtgtgtctgtgatgactgtcagcac1260aacacagaaggacagttgtccagaggtgagttcctgcaatcgggagctgtcagcac1380cttcctgccaactcagtgaccttctgcaactgttcctgccatcgaggacttggagactt1380cttcctgccaactaggacttgtcaggagctgttactggactcggagactttggagactt1500ggctgtcgcacggcgacttgtgacaggtggtcgtagacctatccggagactac1500ggcgaaccaggtagagaacagacattagacgattgtgggatctgcactctatcacactcagg1680aaattgcgaatgtaaggaacagacattagacgcaggggtttctgcactcttcgtacactctgtaggacacaacattatatccagaacatggaaccttggccaaactaaaaggtactactgtccaacctgttgaggacaacattatatccagaacatggaaccttctttggaacaggcattataacaggacttgagaaattttttaa	actocagtoc	ttttccatcc	actogaagga	aggttattt	caeaacttta	tcaccacaca	-
CCCLCACACALL TRACACACACACACACACACACACACACACACACACACAC	acgatacaga	gaacccttac	agtgccaaag	ttcaggagca	actasaasta	accaecates	
cctcaacatt ttacacata tgccatctat gatttcattg tcaagggcag ctgcttctgc aatggccacg ctgatcaatg catactgtt catggccacg catggccacgag cagcccagga 1020 cagcacttgt gcccagga ccccaggag cagccactgc catggcactgt gttacacatg gaagtgtatg tgtacacatg cacatgcacg caccacgagag cagccactgc 1080 cagcactgtg ggaggacac catgaccgg cagtgatgg caacacaggag cagccactgc 1080 gccacacacga agggagaccg ctgcaagtg catggaggag cagctgatgg caaaacgggg 1140 gctccaacg agggagaccg ctgcaagtg cagtgaggag cagctattgg caaaacgggg cagccacgag cagccactga tcattcacacacacacacagaag gacagtattg ccaagaggt aatggggggt ctgttgatga ctgcagacac 1260 gccttctcag ctccagatg ctgcaaaccg ttgtccacac tcattcagac ctccattcagac cccacagaagg ggcaactgat tgacagggt caagcaggtt tctatcatgag atcacgttgt 1380 ctcctctcacac actagtgac cttctcacgac cacagaagggag tgacagtatg ggaggagacacacacacacacacacacacacacacaca	cacatacaac	toctoaaaco	acagtettet	ccctgtcaga	gaaatgacct	Gaacaacaa	
acttyccacz cycletacatg gaagtytatg tyttagcaca accacagoagy cagocactycy cocceptata castsaccyg cattygagagy cagocactycy cocceptata castsaccyg cattygagy cagocactycy cocceptata cattyaccyg agygagy cagogagy cagogagy cagogactycy cocceptataty gygagycac cagogagy attygaaty cycletacygy gygagycac agygaatcy agygagygy cycletacygy gygagycac agygaatcy agygagygy tottygagygy cottoctygac cycletocygy accapagygy accapagygy accapagygy accapagygy cycletocygy accapagygy tycoletocygy accapagygy accapagy accapagygy accapagygy accapagygy accapagygy accapagygy accapagy	cctcaacatt	ttacacacta	toccatctat	gatttcattg	tcaagggcag	ctacttctac	-
acattecacatggtecatgggaagtgtatgtgttageacaacacageaggcagecactge1080cagcactgttgccccgttatacaatgaceggcattgagaaggcagactgatggcaaaacggggg1140geteccaaggagtgeagacactgeaagtgtaatgggeatgctgatacetgtecattetgac1200gttaatgtgggaagcattagggaategtaatgggeatgtetgtatgacettectgac1260accaacagaaggacagtattgccagaggtgcaagccaggettettetetgacceteggaga1320cettectgeaacteagtgacttgcaaaccgtgttectgecatcagtgagtetgaagetg1440ggggtggaggaggggcgacgttgtgcagggagctgtgacectatecacggaggtegaaggget1560ggctgtegacacatagaactggaatcatgaagatcetgactecgtecegtgacaataag1680aaatgcgaatgtaaagaacagacattagaagattectgacctgcactccacatcaaggt1860aaatgtgaagttaaaaataagattttatcaggctacatgaggtattecatgttgaagac1920tttggaataccttgtagacatggcacactaggacataggaaactaaggcacactaggacattataaaaactaa1920tttggaatacctttgtagaaggacattgagaaaagttttecagaacataa1920tttggaatacctttgtagaagagacattagagaaaagtcattaatggacattatagaaaagtcattaatgaaatttttaaagaaaaggagtcaagtagaagaagacattaatgaaaagtcattaatgaaattttttaa2040aagaagagtgcaagaagaactttattttaagagaaactgaattttttaa2220taaatacagt <td< td=""><td>aatggccacg</td><td>ctgatcaatg</td><td>catacctott</td><td>catggcttca</td><td>gacctgtcaa</td><td>adccccada</td><td></td></td<>	aatggccacg	ctgatcaatg	catacctott	catggcttca	gacctgtcaa	adccccada	
cagcactgtg ccccgttata caatgaccgg ccatgggagg cagctgatgg caaaacgggg 1140 gctccaaacg agtgcagaac ctgcaagtgt aatgggcatg ctgtatactg tcacttcgac 1200 gttaatgtgt gggaggcatc agggaatcgt agtggtggtg tctgtagatga ctgtcagcac 1260 acacacagaag gacagtattg ccagaggtgc aggcaggct tctatcgtga cctgcaggag 1320 cccttctcag ctccagatgc cttctggac ccagagggt tctatcgtga cctgcaggag 1320 cccttctcag ctccagatgc cttctggac ccagagggt ttgtactgag atcaggtgt gggggtggag gggggtggag gggggggggg	acattccaca	tggtccatgg	gaagtgtatg	tgttagcaca	acacagcagg	cadccactdc	
gctcccaacg agtgcagaac ctgcaagtgt aatgggcatg tctgatacctg tcactccgac 1200 gttaatgtgt gggagcatc agggaatcgt agtggtggtg tctgtagtag ctgcagaag gacagtattg ccagaggggc aagccaggct tctatctgtg atcacgtag ctccagtagg ctccagaagggggggggg	cagcactgtg	ccccgttata	caatgaccgg	ccatgggagg	cagctgatgg	caaaacaaaa	
gttaatgtgt gggaggcatc agggaatcgt aggggtgtg aagccaggct tctatcgtga ctgcaggag 1320 cccttctcag ctccagatgc ttgcaaaccg tgttcctgcc atccagtagg atcagctgtc 1380 cttcctcgca actcagtgac cttctgcgac actcagtgag actgggggtgggggggggg	gctcccaacg	agtgcagaac	ctgcaagtgt	aatgggcatg	ctgatacctg	tcacttcgac	
acacagaag gacagtattg ccagagtgc aagccaggct tctategtga cctgeggaga 1320 ccctteteag ctccagatgc ttgcaaaccg tgtteetgcc atccagtagg atcagetgte 1380 ctteetgeac actcagtgac cttetgegac cccagcaatg gtgactgcec ttgcaagcct 1440 gggtgggag ggcgacgttg tgacagggg atgggggggggg	gttaatgtgt	gggaggcatc	agggaatcgt	agtggtggtg	tctgtgatga	ctgtcagcac	
CCCLTCCTGGGCTCCAGGATGGCTGCAGGATGGGCTGCAGGATGGG <t< td=""><td>aacacagaag</td><td>gacagtattg</td><td>ccagaggtgc</td><td>aagccaggct</td><td>tctatcgtga</td><td>cctgcggaga</td><td></td></t<>	aacacagaag	gacagtattg	ccagaggtgc	aagccaggct	tctatcgtga	cctgcggaga	
cttcctgcca actcagtgac cttctgcgac cccagcaatg gtgactgccc ttgcaagcct 1440 ggggtggcag ggcgacgttg tgacaggtgc atggtgggat actggggctt cggagactat 1500 gcctgccacacag acatagacgg ttgcagggac tgtgacccta tcaccggaga ctgcatcagc 1560 agccacacag acatagacgg gtatcatgaa gttcctgact tccgtcccgt gcacaataaag 1620 aaatgcgaat gtaaggaaca gacattagaa aatgccaaagg cattcttgga aatgaaatat 1740 tcatatggc taaaaataaa gattttatca gctcatgata aaggtactca tgttgaggtc 1880 aatgtgaaga ttaaaaaaggt cttaaaatct accaaactga aggatgccc tgttgaggtc 1880 aacattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 aagaaaggtct aagatgagat gacactacat ggacacctc tctgagaga aaggacacta tcttgtgaatac 2040 aagaaggaggc aagtaggact tataagaaca gcacataatga agacttcaaaa 2160 caatatagaaa tgtttgaatg catgggtct tatactaatt accagtatac tctgaaacccc acacaggcccccccccc	cccttctcag	ctccagatgc	ttgcaaaccg	tgttcctgcc	atccagtagg	atcagctgtc	
ggggtggcag ggcgacgttg tgacaggtgc atggtgggat actgggggtt cggagactat 1500 ggctgtcgac catgtgactg tgcaggagac tgtgacccta tcaccggaga ctgcatcagc 1560 agccacacag acatagactg gtatcatgaa gttcctgact tccgtcccgt gcacaataag 1620 agcgaaccag cctggggtg ggaggatgcg caggggtttt ctgcacttct acactcaggt 1680 aatgcgaac gtaaggaaca gacattagga aatgccaaagg cattctgtgg aatgaaataat 1740 atcattgtgc taaaaataaa gattttatca gctcatgata aggatactca tgttgaggtc 1800 atgttggaaga ttaaaaaggt ctttaaaatct accaaactga aggataccat tgttgaggtc 1800 agaattatat ccagaatcat ggacggacag aggatgacact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaaggtgc aagtagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagaggagggc aagtagaac gacctcagac aggaaactgg atattttaa agtgccaaaa 2100 aagagaggggacga tttattcat gaaggaaactgg aatttttaa agtgccaaaa 2160 aagatacacc tcagaacct taaccaccct acactgatat ctgtcaaccc acactgatat ctggaacccc acactgatat ctgtaacccc acactgatat ctgtttcta 2280 tcagacctgaaccccc acactgatacccccc acactgatacccccccccc	cttcctgcca	actcagtgac	cttctgcgac	cccagcaatg	gtgactgccc	ttgcaagcct	1440
agccacacag acatagacty gratcatgaa gractctagact teegteeegt geacaataag 1620 agcggaaccag Cetgggatty ggaggatgeg caggggtttt etgeacttet acacteaggt 1680 aaatgegaat gtaaggaaca gacattagga aatgecaagg cattetgtgg aatgaaatat 1740 teatatgtge taaaaataaa gatttatea geteatgata aaggtactea tgttgaggte 1800 aatgtgaaga ttaaaaaggt ettaaaatet accaaactga aggattteeg agggaaageg 1860 acattatat eegaaaccat ggacggacag aggatgeact tgtecaatee teaateetgg 1920 tettggaatae etgttgagga tataagaaca ggacattagaga tataagaaca ggeaaactaa ttgtgaatat 1980 gaaaagggaggaaggaggaagggaaaccte teettggaaga aaagteatgg atatttaaa 2040 aagaagagggagaaccte teettggaaga aacattatagaaa tgtttgaaga gaceteagac aggaaacetg aatttttaaa aggacacaaa aggaaacctg eatgggtet tateetaatt ateetatte ateettetg gacecatgggact tateetaatgaaa teettetteg gacecatggaaccet tateetaate ateetteteg gacecatggt 2220 tataatagaac tettatteea teagggacga tateetaate agaaaccee teagggactgat tettateaa agaaaccee teagggactgat tettateaa agaaaccee teagggacgact tateetaatte ateetteeg gacecatgtt 2220 tataatacagt tettatteea teaggacgac taacetaate agaaaccee teeaggacgact teagaaacce teagggacgact teagaaaccee teaggacgact teagaatee teaggacgac teagaaccee teeagetea getattace 2340 taacgtttga tgetacetaa aggacgeete tegggacee teggaaaccea accetaate 2400 teagaaaagge gettgttaaa ggaceetee eatgtgtata tagtattee teggaaaccee tegggtgttaaaa ggaceetee eatgtgtata tagtattee teggaaaccee teggaaaccee eatgtgtaa aggacgeete teggaaaccee teggaaaccee eatgtgtata tagtattee teggaaaccee teggaaaccee eatgtgtata tagtattee teggaaaccee teggaaaccee eatgtgtata aggaceetee eatgtgtata aagaaggaaa agggtgtgtg 2640 agaaagcea eatttaaacce tegaaaccee tegaaaccee eatacaacaggaaaggaaaccee eatacaacaggaaaggaaaccee eatacaacaccee eatacacacee eatacacacacee eatacacacee eatacacacacee eatacacacacee eatacacacacee eatacacacacee eatacacacacacacacacacacacacacacacacacaca	ggggtggcag	ggcgacgttg	tgacaggtgc	atggtgggat	actggggctt	cggagactat	1500
agegaaccag cetgggagtg ggaggatgeg caggggtttt ctgcacttet acactcaggt 1680 aaatgcgaat gtaaggaaca gacattagga aatgccaagg cattctgtgg aatgaaatat 1740 tcatatgtgc taaaaataaa gattttatca getcatgata aaggtactca tgttgaggtc 1800 aatgtgaaga ttaaaaaggt cttaaaatct accaaactga agattttccg agggaaagcg 1860 acattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaaggaggtg aagtagcatt gtccagacct ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagaggaggtg aagtagcatt aggacgaaaa gacctcaaga aggatgcact atgtccaaaa 2100 cacaaacttt agaagcagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 aagaggactgat tctgaaatc ttattcat gaagagaaat tactaatta accactgatat tctgaaatc cagaaaaccgc taacaccct acactgatat ctgaaatcc ataagggcag tacctctaac agaaaacccc acactgatat ctgacattag 2280 tacggtttgaatca tgcacattaa ggacctca aggaacccct tccagettca gcacattatag 2340 ctacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactaa 2460 ctacgaaaaggt gtttgtaaa ggacctcc catgtgtata tagtattcc ttgtgtgta cgcatgaac ttggtcgta 2520 ggtgttaaat ggacctcc catgtgtata tagtattcc ttgttgaaa aggggtgtgg 2640 agaaagccta ctgaagcag agcactgcc ctacatgtgg acaaaagga accataaaa 2700 gaagattggc tatttaacc ttgaaaccc tacaaaggaaggaaaccgaaaggaagaaggaaagga	ggctgtcgac	catgtgactg	tgcagggagc	tgtgacccta	tcaccggaga	ctgcatcagc	1560
aaatgcgaat gtaaggaaca gacattagga aatgccaagg cattctgtgg aatgaaatat tcatatgtgc taaaaataaa gattttatca gctcatgata aaggtactca tgttgaggtc 1800 aatgtgaaga ttaaaaaggt cttaaaatct accaaactga aggattttccg agggaaagcg 1860 acattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaagcttt gtccagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagagaggtgc aagtagcact aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agaagcagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tttatttaat gaaggacat tatactaatt acctttttg gacccatgtt tatattaatacagt tttatttcat gaagagaaat gaaaacccct acacctgatat ctgtttcta 2280 taaggactgat tctgaaatcc ttaactatta agaatattt accgtttga tgcacattaag gaccttctaac aaggacgcct tccagctca gctatgttac 2400 tacgtttga tgctacttaa gaccttctc ttgtgtgta cgcatgaact ttggtctgta 2520 ggtgttaaat ggacctctc catgtgtaa tagtatttcc ttgtgtgta aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcag agcactgca ctacatgtgg accaaaagggaagtgaaggaagtgaagcta ctgaagaag agcactgca ctacatgtgg acaaaagggaagcg agaaggaaggaaggaaggaagg	agccacacag	acatagactg	gtatcatgaa	gttcctgact	tccgtcccgt	gcacaataag	1620
tcatatgtgc taaaaataaa gattttatca gctcatgata aaggtactca tgttgaggtc 1800 aatgtgaaga ttaaaaaggt cttaaaatct accaaactga aggatttccg agggaaagcg 1860 acattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaagcttt gtccagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagagaggtgc aagtaggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agaagcaagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtct tatctaattt atctcttctg gacccatgtt 2220 tttaattcat tctgaaatc ttaactatta agaacaccct acactgatatcttcta tctgaaatcc ttaactatta agaacaccct acactgatatcttcta 2340 cagaaaaggt gtttgtaaa gagccttctc ttgtgtgta cgcatgaac tccgacactat 2460 cagaaaaggt gtttgtaaa gagccttct ttgtgtgta cgcatgaac ttggtcttacta 2520 ggtgttaaat ggacctctc catgtgtgaa cgtttggtga agaaaccca ttggtgtgta accacactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaa aggggtgtgg accacacgaaggaagtgaagcct ctgaagacgc tacacactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa aggggtgtgg accacacgaaggaagtgaagt	agcgaaccag	cctgggagtg	ggaggatgcg	caggggtttt	ctgcacttct	acactcaggt	1680
aatgtgaaga ttaaaaaggt cttaaaatct accaaactga aggatttccg agggaaagcg 1860 aacattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaagcttt gtccagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagagagtgc aagtagcatt aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agaacagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtct tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttattcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaattc ttaactatta agaatattt aatagcagca tgacatttag 2340 cagaaaaggt gtttgtaaa gagccttctc ttgtgtgta ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtgata tagtatttcc ttgtgtgta aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcag agcactgcca ctacatgtg gtgaagatg aaccagaaag 2760	aaatgcgaat	gtaaggaaca	gacattagga	aatgccaagg	cattctgtgg	aatgaaatat	1740
aacattatat ccagaatcat ggacggacag aggatgcact tgtccaatcc tcaatcctgg 1920 tttggaatac cttgtagcag gacatgagga tataagaaca ggcaaactaa ttgtgaatat 1980 gaaaagcttt gtccagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagagagtgc aagtagcatt aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agagcaagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtctt tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaatcc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcag tacttacta gagaaactag gtgaagatgc aacaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc ttgaatactt gagaaactag gtgaagatgc aaccagaaag 2760	catatgtgc	taaaaataaa	gattttatca	gctcatgata	aaggtactca	tgttgaggtc	
tittggaatac citgtagcag gacatgagga tataagaaca ggcaaactaa tigtgaatat gaaaagctit gtccagcact ggaaacctic tcttggaaga aaagtcatgg atatittaaa 2040 aagagagtgc aagtagcatt aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaactit agagcaagaa gacctcagac aggaaactgg aatitittaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtcti tatctaatit atctctictg gacccatgti 2220 taaatacagt titatitcat gaagagaaat gaaaacccci acactgatat ctgttitcta 2280 tgggactgat tctgaaatic ttaactatia agaatatiti aatagcagca tgacattag 2340 cagtaatcca taagggcag tacctctaac aaggacgcci tccagctica gctatgttac 2400 tacgittga tgctactaa agtaatgaat gacgitttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgitaaa gagcctictc tigtgtgta cgcatgaact tiggitcigta 2520 ggtgitaaat gagccctcc catgitgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcag agcactgcca ctacatgtgg acaaaaggtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatactt gagaaactag gtgaagatgc aaccagaaag 2760	aacgcgaaga	ctaaaaaggt	cttaaaatct	accaaactga	agattttccg	agggaaagcg	
gaaaagcttt gtccagcact ggaaaccttc tcttggaaga aaagtcatgg atattttaaa 2040 aagagagtgc aagtagcatt aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agagcaagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtctt tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttc ttgtataaag cactttacta 2580 cctaccactt gtgttgtaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc tacttacta gagaaactag gtgaagatgc aaccagaaag 2760	tttggaatag	cttatagaac	ggacggacag	aggatgcact	tgtccaatcc	tcaatcctgg	
aagagagtgc aagtagcatt aagatggata gcacataatg gcacttggct atgtccaaaa 2100 cacaaacttt agagcaagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtctt tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgttttcta 2280 tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttc ttgtataaag cactttacta 2580 cctaccactt gtgttgtaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc tacttacta gagaaactag gtgaagatgc aaccagaaag 2760	gaaaagettt	atagagaga	gacatgagga	tataagaaca	ggcaaactaa	ttgtgaatat	
cacaaacttt agagcaagaa gacctcagac aggaaactgg aatttttaa agtgccaaaa 2160 catatagaaa tgtttgaatg catgggtctt tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttc ttgtataaag cactttacta 2580 cctaccactt gtgttgtaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc tacttacta gagaaactag gtgaagatgc aaccagaaag 2760	aagagagtgc	aagtaggatt	aggatagette	gcacataatg	adagtcatgg	atattttaaa	
catatagaaa tgtttgaatg catgggtctt tatctaattt atctcttctg gacccatgtt 2220 taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	cacaaacttt	adadcaaca	aagatggata	gcacataatg	geacttgget	atgtccaaaa	
taaatacagt tttatttcat gaagagaaat gaaaacccct acactgatat ctgtttcta 2280 tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	catatagaaa	tatttaata	catogotott	tatctaattt	atctctttaa	aytyccaaaa	
tgggactgat tctgaaattc ttaactatta agaatatttt aatagcagca tgacatttag 2340 cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	taaatacagt	tttatttcat	daadadaaat	datacaaccc	acectette	gacccatgtt	
cagtaatcca ttaagggcag tacctctaac aaggacgcct tccagcttca gctatgttac 2400 ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	tgggactgat	tctgaaattc	ttaactatta	agaatattt	aataggagg	tracatttar	
ttacgtttga tgctacttaa agtaatgaat gacgttttaa ggaatcccta accctactat 2460 cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	cagtaatcca	ttaagggcag	tacctctaac	aaggacgcct	tccagcttca	actatattaa	
Cagaaaaggt gtttgttaaa gagccttctc ttgtgtgtta cgcatgaact ttggtctgta 2520 ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	ttacgtttga	toctacttaa	agtaatgaat	gacgttttaa	ggaatcccta	accetactat	
ggtgttaaat ggaccctctc catgtgtata tagtatttcc ttgtataaag cactttacta 2580 cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	cagaaaaggt	gtttgttaaa	gageettete	ttatatatta	CGCatgaact	ttaatctata	
cctaccactt gtgttgtgaa cgtttggtga ctgctgttga aagaaggaaa agggtgtgtg 2640 agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	ggtgttaaat	ggaccctctc	catqtqtata	tagtatttcc	ttgtataaag	cactttacta	
agaaagccta ctgaagcagc agcactgcca ctacatgtgg acaaaagtga acatataaaa 2700 gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	cctaccactt	gtgttgtgaa	cgtttggtga	ctgctgttga	aagaaggaaa	agggtgtgtata	
gaagttgtgc tatttaactc tgaatacttg gagaaactag gtgaagatgc aaccagaaag 2760	agaaagccta	ctgaagcagc	agcactgcca	ctacatgtgg	acaaaagtga	acatataaaa	
gagaatatgt atgcgtgaag tctcagcttt gagctggagg ctagattcca agatgacagc 2820	gaagttgtgc	tatttaactc	tgaatacttg	gagaaactag	gtgaagatgc	aaccagaaag	
	gagaatatgt	atgcgtgaag	tctcagcttt	gagctggagg	ctagattcca	agatgacagc	

£ 1.

```
catgatgaaa ctttttaaaa aactaaacca gaagagactt taaaataaga gaaagaaatc
                                                                      2880
  ataaatgtag acatatgctt ggctaaaggg gaaatggact ttaaatttta aagagctcat
                                                                      2940
  ttgcaatgca cttgtataca cttcaaaaat tattgtagac acagaatttg ttatatttt
  gtgcttagta tttaaacctg aacattgaaa cagttttcct ccttgtcttt cttaacagta
                                                                      3000
                                                                      3060
  atagtcatta tatttacctg ttttttaaca caatgtatgt gatagtcaaa aaatcacagt
  ttttcattat tattcatctt ctgtacccac gcataaccac tatacatagt ttcttttgta
                                                                     3120
 cttgaatata caaaacatga acacagtgcc atatgaataa tttcacatac agaacctttt
                                                                     3180
 tttctctgaa gtcctgtgga cttgcaaata tatatatat ttgctttgtt aatttgtttt
                                                                     3240
                                                                     3300
 3360
 ggggggccc gtacccaaa
                                                                     3379
 <210> 28
 <211> 2006
 <212> DNA
 <213> Homo sapiens
 <400> 28
 cccacgcgtc cgcggacgcg tgggattttc tcttaagaaa aaaaaaggaa gacgttttaa
                                                                       60
 taggaaaaga aaaacaagaa gcagtagcag ctaggtaatt taaacgtctt ccttttttt
 tottaagaga aaatggaaca tttaggttaa atgtotttaa attttaccac ttaacaacac
                                                                      120
 tacatgccca taaaatatat ccagtcagta ctgtatttta aaatcccttg aaatgatgat
                                                                      180
 atcagggtta aaattacttg tattgtttct gaagtttgct cctgaaaact actgtttgag
                                                                      240
 cactgaaacg ttacaaatgc ctaataggca tttgagactg agcaaggcta cttgttatct
                                                                      300
 catgaaatgc ctgttgccga gttattttga atagaaatat tttaaagtat caaaagcaga
                                                                     360
                                                                      420
 tottagttta agggagtttg gaaaaggaat tatatttete ttttteetga ttetgtacte
 aacaagtett gatggaatta aaataetetg etttattetg gtgageetge tagetaatat
                                                                     480
aagtattgga caggtaataa tttgtcatct ttaatattag taaaatgaat taagatatta
                                                                     540
                                                                     600
taggattaaa cataatttta tacggttagt actttattgg ccgacctaaa tttatagcgt
                                                                     660
gtggaaattg agaaaaatga agaaacagga cagatatatg atgaattaaa aatatatat
                                                                     720
ggtcaatttt ggtctgaaat ccctgaggtg tttttaacct gctacactaa tttgtacact
aatttatttc tttagtctag aaatagtaaa ttgtttgcaa gtcactaata atcattagat
                                                                     780
aaattatttt cttggccata gccgataatt ttgtaatcag tactaagtgt atacgtattt
                                                                     840
                                                                     900
ttgccacttt ttcctcagat gattaaagta agtcaacagc ttattttagg aaactggaaa
agtaataggg aaagagattt cactatttgc ttcatcagtg gtaggggggc ggtgactgca
                                                                     960
actgtgttag cagaaattca cagagaatgg ggatttaagg ttagcagaga aacttggaaa
                                                                    1020
gttctgtgtt aggatcttgc tggcagaatt aactttttgc aaaagtttta tacacagata
                                                                    1080
tttgtattaa atttggagcc atagtcagaa gactcagatc ataattggct tattttcta
                                                                    1140
                                                                    1200
tttccgtaac tattgtaatt tccacttttg taataatttt gatttaaaat ataaatttat
ttatttattt ttttaatagt caaaaatctt tgctgttgta gtctgcaacc tctaaaatga
                                                                    1260
ttgtgttgct tttaggattg atcagaagaa acactccaaa aattgagatg aaatgttggt
                                                                    1320
                                                                    1380
gcagccagtt ataagtaata tagttaacaa gcaaaaaaag tgctgccacc ttttatgatg
attitctaaa tggagaaaca tttggctgca tccacataga cctttatgtt ttgttttcag
                                                                    1440
ttgaaaactt gcctcctttg gcaacattcg taaatgaagc agaattttt tttctcttt
                                                                    1500
ttccaaatat gttagttttg ttcttgtaag atgtatcatg ggtattggtg ctgtgtaatg
                                                                   1560
aacaacgaat tttaattagc atgtggttca gaatatacaa tgttaggttt ttaaaaagta
                                                                   1620
                                                                   1680
tottgatggt tottttctat ttataatttc agactttcat aaagtgtacc aagaatttca
taaatttgtt ttcagtgaac tgctttttgc tatggtaggt cattaaacac agcacttact
                                                                   1740
                                                                   1800
cttaaaaatg aaaatttctg atcatctagg atattgacac atttcaattt gcagtgtctt
tttgactgga tatattaacg ttcctctgaa tggcattgat agatggttca gaagagaaac
                                                                   1860
                                                                   1920
aaagggcggc cgctctagag gatccc
                                                                   1980
                                                                   2006
```

<210> 29

<211> 3070

<212> DNA

<213> Homo sapiens

gttggagct tagacettcg aagtgaccta tagaaggtac gcctgcaggt accggtccgg 60 aattcccggg tcgaccacg cgtccgcaac gtatgaggag tctgatttet tcacatcgt 120 gtcaacattt gttgttgttg ttgttgttgt tgtttttaat tttagattac tcatgatagtg ttgaacattt gttgttgttg ttgttgttgt tgtttttaat tttagattac tcatgatagtg ttttcatgtg cttgttgcc attcgtatat cttcttcaat gactagtgat gttgaacatc 240 ctttcatgtg cttgttgcc attcgtatat cttcttgag aaatgctat taagtcctt 300 gcccaattc aaattggctt gtcttttgt tgagtataa ccattgaatt ttaattcgt 360 taaaacactc catgagtaca tgtagcatga actcaataac tattagctaa taatgctagt tgaaattatt ttgcccaatag tggcctgaat tggaaggaaa tgcctagtt ttacaagta ggagtggaga tgtaacaaaa cctgtattta aggccaataa agatttgact ctgtagtgt 540 ggaatggaga tgtaacaaaa cctgtattta aggccaataa agatttgact ctgtagtgt 540 gaatcagttta tctttgacga aaatcagtaa attatacaatg tttaaattaat tcctaaggat 600 aatcagttta tctttggca aaatcagtaa attatacaatg tttaaattaat tcctaaggat 600 aatcagttta tctttgaga aaatcagtaa attatacaatg tttaaattaat tcctaaggaa 600 aatcagttta tctttaggg tgttcaaa aggattaga aggattagat cattcttet cattctgccc 720 tgttatgcga catttgctat aaagtcaca atgagaattt tttgaaacct gggtcagaga 780 agcctacttt ttaatttaa taagtatgtg attgatgtat tatcattagg ttatgatgtat tatcattagat 120 tgcaatctgg tttttagggg tgttcaaa agaatagtg agaatagtg agaataggg agaataggg attgaggaa accaactg 120 atttagaga acaacagca tatgtaagtc ccgttcatt atattggtga tttggagta gggaacagga 102 tttaagcaga acaacagca tatgtaagtc tcattattagt ttttttggagt aggaatagga atatttccaag gaactggg tgtagaactg tatcgatgga tcatgagaa atatttcaa ggagattat tcctgag gaattggg tgtagaacaga acaacagca 1020 tttaagcag accaacag aggattgg tccaaatac tcctgtagag ctctctctcc ttagcgatt atgcattcag gaattggg tgtagaatt tttgagagc tcctctccc tcaactgag tctttcggg tgggtacaaa gagttgtgt tttaaaaattgg cagttagaa acaactgtaa acaactgtaa acaactgtaa acaactgtaa acaactggaa acaactggaa caactggaa acaactggaa acaactggaa acaactgtaa acaactg	4400× 20						
acticacing tegaceacy egicegeacy gracings cities to testated teacatect gracings of the gracing to the state of	<400> 29				~~~*	3.55555t 5555	60
gtcaacattt gttgttgttg tigttgttgt tgttttaat tttagcattc ctagtagag 180 tgaagtggta tctcgtgctt tttatttgca tttctctaat gactagtgat gttgaacatc ttttcatgtg cttgttggcc attogtatat cttctttagag aaatgctaat taagtccttt 300 gcccaatttc aaattggctt gtctttttgt tgagttataa ccattgaatt taagtcctt 420 taaaacactc cattgagtaca tgtagcatgc actoaataac tattagctaat actatagctt gtaattattt ttgccacatg tggcctgaat tggaaggaaa ttgcctagtt tttccaagta ggagtggaga tgtaacaaaa cctgtaatta aggccaataa agatttgact ctgtscagtt tggaattggaga tgtaacaaaa cctgtaatta aggccaataa agatttgact ctgtscagtg tgatactttg cttttaattg atgccccaa atcaagaaat tagaaacgtg tatacataag 600 aatcagttta tctttggcag aaatcagtat attacaatg tttaattaat tcctaagaat aagcctacttt ttatatttaa taagttagtg aggattagat cattttett cattctgccc tgttatgcga catttgctat aaatgccaca atgagaatt tttgaaacct gggtcaaga agcctacttt ttatatttaa taagtagtg atgstgat attacattggt tatacataag agcctacttt ttatatttaa taagtagtg atgstgat accaagtag agcattagtg tatacatag agctactttg ttttatagggg tgtcaaaag aagatagag agaataggtg atccaaaga agctacttt tgggatta ggggggggggggggggggggg							
type technique tetratique attentique attenti		-					
ctticatggg cttgtiggce attogtatat cttittgg aaatgtciat taagtccttt 3300 geccaatite aaattggett gictititgt tgagtiataa ccattgaati tiaattgcig 360 taaaacactc catgagtaca tgtagcatga catcaataac tattagctaa tactatagtt 420 giaattatit tigccacatg tggcctgaat tggacagaaa ttgcctagti titiccaagta 480 ggagtggag tgtaacaaaaa cctgattita aggccaataa agattigact ctgtcagti 540 gaatcagtit tctittaatig atgccccaa atcaagaaat tagaaacgtg tatacataag 660 aatcagtita tctitiggcag aaatcagtat attacaag titiaattaat tctitagcag aaatcagtat attacaag taattacatag titiaattagt catcititit cattcigccc 720 tgttatggga cattigctat agattggg agattgaga catcitititititicatat cattiggcag aaatcagtat attacaag titiaattaat tcctaagaaa 780 gagcctacttit titiattata taagtggg agattaggt attacatag gggtcagaga 880 gagctacttit titiattata aaatgccaca atgagaatti titigaaacct gggtcaagag 880 gagctacttitititititiaa gagtagtgg giticaaaag agataggtg actacatagtg attacatagtg attatactag gattiggagti titiaggagtg acacactig aactgatgaa tatatictac 960 caggititata ggtagaatat ccgtiticatt atatcggga titiggagtga gacaagaagci 1020 titiaagcaga acacacgcca tatgtaagtc taaattitigt tattitictit gaataagca 1080 gaaactgti atgcatactig getticaag gacactggga titiggagtga titiggagtga gacagtagci 1020 titiaagcaga acacacgcca tatgtaagatc taaattitigt tattitictit gaataagcaa 1020 titiaagcaga acacacgcca tatgtagatc tectgaatac titiggagtga titiggagtgaga gagaatagtga titiggagtga attitiaa aggacttata toctattititi toctiticaaa agattititi gctgataca agattagaa aaattaagga catgacagaa ataaatggg catgagaa aaatagagaa titiggagtgaga gagagagaa taattaga gagaataga aaattaaga gagaataga titiggagtgaga gagaatagaga titiggagtgaga gagaatagaga titiggagtgaga aaaaaaaaa aaaaaaaaa aaaaaaaaaa							
gcccaattt daattgctt gtettttgt tgagttataa ccattgaatt ttaattgctg taaaacactc catgagtaca tgtagcatgc actcaataac tattagctaa tactatagtt 420 gtaattatt ttgccacatg tggcctgaat tggaaggaaa ttgcctagtt tttccaagta 480 ggagtggaga tgtaacaaaa cctgtaatta aggccaataa agatttgact ctgtcagtgt 540 tgatactttg cttttaattg atgccccaa atcaaggaaat tagaaacggt tatacaataag 600 aatcagttta tctttggcag aaatcagtaa attatcaatg tttaattaat tcctaaggaa 660 taatggcaat aggccctac gttttagctg agattaggt catcttttt cattctgccc 720 tgttattgg acattgcta aaatgcacaa atgagaattt tttgaaacct gggtcaagga 780 agoctacttt ttatattaa taagtagtg attgatgat tatcaattgg tttaaggaa 780 agoctacttt ttatattaa taagtagtg atgagaattg tttagatgg tttaaggag 780 agoctacttt ttatattaa taagtagtgg acacacctg ggagaaggag agacaaggag agaatgagg agaatgagg agaatgggg agaatgggg agaatgggg agaatgggg agaatgggg agaatgggg 1020 atttaaggg tgtcaaaag agaatggggg acacactgg gaacagggggggggg							
taaaacactc catgagtaca tgtagcatgc actoataac tattagctaa tattatatt tgtacatatt ttgccaatt tggactgaat tggaaggaa ttgccaatta tggaaggaa ttgccaatta tggaaggaa ttgccaatta tggaaggaa ttgccaatta tggaaggaa ttgccaatta tggaaggaa ttgccaatta ggaaggaagaa ttgccaatta ggaaggaagaa ttgccaatta ggaaggaagaa ttgtaacaaaa cctgtattta aggccaatta agatttgact ctgtcaagat 600 aatcagtta tctttagcag aaatcagta attatcaatg ttaattaata tcctaagaat 660 taatggcaat aggccacta gtttagcag agattagat catctttctt cattctgcc 720 tgttaatgcga cattigctat aaatgccaac atggaattt tttaattaat tcctaaggat 780 aggccacttt tttaattataa taagtatgtg attagtgat tatcaattagg ggtcaagaa 780 aggccacttt tttaattataa taagtatgtg attagtgat tatcaattagg ggtcaagaag 780 aggctactttg tttatgggg tgttcaaaaa gagatagatg agaataggtg atccaaagta attatcac 960 caaggtttata ggtagatag cgtttcaat atatcactggtga tttggaagtag acaacagcac tatgtaagtc taaatttggt tattgtttt gaataagca 1020 tttaagcaga acaacagcac tatgtaagtc taaatttggt tattgtttt gaataagca 1080 tgggaactg acaacactgt acaacagtgga ttcagaatta tcctgaagac ctctcccc 1140 ctgtgctcc acagctttt tcctgaactc tcctgtctc ttggaagcc ctctcccc 1140 ctgtgctcc acagctttt tcctgaactc tcctgtctc ttggaagcag ccttctccc 1140 ctgtgctca acagctatt tcctgagg ttgtacaaa tgctgcttta gccattaaa acttgtacaa 1260 tcctactaatt gtgaactcta cctgaactca tcctagtgag ttttaaaaaa acatgaaga 1380 atttttaaag cagtgcatgc atgcagtaaa caaaattgga cagtacagaa ataaatggg 1440 aagaacaagct cccactgaag gaatttttat accgtatgaa gtgcttatta accttaaaaa 1380 atttttaaag ggaatagaa agaattattat accgtatgaa gtgcttatta accttaaaaa 1560 gatattttg gtgaacaaa agattttat accgtatgaa gtgcttatta accttacaaa 1560 tcctaataa catcctct taccttagac gaattttat accgtatgaa gtgcttatta accttaaaaa 1560 gatattttat ggaacagaa tttccacaaa aaatcaaga catacaaaaa 1560 gatagaagaaa aggccccct taccacactg gaattttaa accgtatgaa gcgtcactta accttagtgt 1740 aggacagaa ataacacgg gtgaacacaa accttggt tccacagaaa acctctgt gcgaaggaagga accacacaa tgaaggaga accacacaa tccacagaa accacacaa tccacagaa accacacaa accacagaa accacacaa accacagaa accacacaa accacagaa accacacaa accacagaa cccacaaaaa accacacaa accacagaa cccacac							
graattattt ttgccacatg tggcctgaat tggaaggaaa ttgcctagtt tttccaagta 540 ggagtggaga tgtaacaaaa cctgtattta aggccaataa agatttgact ctgtcagtg 540 tgatactttg cttttaattg atgctccaa atcaagaaat tagaaacgg tatacataag 660 aatcagttta tctttgagca gaattagtat attatcaatg tttaattaat tcctaagaat 660 taatggcaat agtgcccta gttttagcta agatttagat catctttctt catcttgccc 720 tgttatgcga catttgctat aaatgccaca atgagaattt tttgaaacct gggtcagaga 780 agcctacttt ttatattaa taagtatggt attgattgat tatcattggt tttagctag 490 atttagaatt ttatacttaga tttgatgatt tatcattggt tttagcga 290 atttagaatt ttatacttag 540 taagtggtga accatcttg ttttagggg tgtcaaaag aagatagatg agaataggtg accagagta 990 atttagaact ttatactccc gttgggactg aaccatcttg accatctgg accagagaa 780 attgagcaga acaacagca tatgtaggt aacaatcttg aactgatgag acaataggcg 780 attgagcaga acaacagcca tatgtaagt taaatttggt tatttcttt gaataagca 1080 tggaaactgt atgcattcag gaactggtgg ttcagaatatc ttctgaaggc ctcttctcc 1140 tggaactct decagtgttg tttgagatgc gacagtgagct 1020 tacctatttt gtgaactctg cctgaactc tcctgttctc ttggcgtgcc tctgtgccta 1200 tacctatttt gtgaactctg cctgaactc tcctgttctc ttggcgtgcc tctgtgccta 1200 tacctatttt gtgaactctg ctggatcaaa tgctgcttt agccattaaa atctgtcaaa 1320 tctgttacat actttcacaa tagcactta cctagtgtgg ttttaaaaat gcacttaaaa 1320 tctgttacat actttcacaa tagcactta cctagtgtgg ttttaaaaat gcacttaaaa 1320 gatattttag gtgaatcaaa gaattttta accgatgtgaa gagacagaa atctaattt tcctcaaa 1320 gatattttag gtgaactaaa agatcattc tctttttta tccttcaaa 1320 gatagaagaaa aggcccctt tcccatgg aagacactt tctttttta tccttcaaa 1560 ttacttcagg gtaatttatg gatagaaa aaatcaagca ttacaacaaaa tgtagaggtg 1620 gatagaatga agatagatga 200 tccttattt tcctaaaaactg 1620 gtgaagaaa aggcccctt tcccagagaa aaatcaagca ttacaacaaaa tgtagaggtg 1620 gtgaagaaga agagcccctt tcccagagaa aaatcaagca ttacaacaaaa 1620 ttactcaaacaaaa aggcccctt gtgaagaaga aagaagaaga aagaagaaga accaccttgtg 200 ttacaaaacaagaaca 1620 ttacaaacaaaa aggccccttaaaaaaa 1620 ttacaaacaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa							
ggagtggaga tgtaacaaaa cctgtattta aggccaataa agatttgact ctgtcaqtgt 540 tgatactttg cttttaattg atgctccaa atcaagaaat tagaaacgtg tatacataag 600 aatcagttta tctttggcag aaatcagtat attatcaatg tttaattaat tcctaagaat 666 taatggcaat agtgctcta gttttaggt gagtttagat catcttctt cattctgcc 720 agataggcacttt ttatattaa taagtacaca atgagaattt tttgaaact gggtagagag 780 agcctacttt ttatattaa taagtagtg attgatgat tttgaagact gaccaacgaga 780 agcctacttt ttatattaa taagtagtg attgatgat ttatcattggt ttatggagg 780 agcctacttg tttatgggg tgttcaaaag aagatagatg agaataggtg atccaaagta 900 atttgagatt ttatcggg tgttcaaaag aagatagatg aacaactgg accaactgggaatt ttatggagatg accaactgt accaacgagta attgagaat ttatctcca gtgggactg aaccaactgt accaacgaga acaacacgca tatgtgaagt ttatggtgat ttattggtga daccaacgagt 1020 tttaagagag acaacacgca tatgtgaagt ttacagatact ttctggaggg gacaggaggt 1020 tttaagagag acaacgagtg tccgaagg ctcccaag cagttggtga cagttggca 1020 ttgggaacg acaggtggg tgccaaga cagttggga cagttggca 1020 tccgtgacca cagggtgggg tgccaaga cagttggga cagttggca 1020 tccgtgacca acgggtgggggggggggggggggggggg							
Egatacittg cittitaettg atgctccaa atcaagaaat tagaaacgtg tatacataag 600 aatcagttat tcittiggcag aatcagtat attacaarg titaataaat tcctaagaat 660 taatggcaat atgtgctcta gittitagctg agattiagat catcittict cattitigcc 720 tgittatgcga cattitgctat aaatgccaca atgagaatti titgaaacct giggtcagaga 780 agcctactit titaattaat taagtagtg attgatgat tatacatagg tatacaaagta 900 atttagaatt titatcatca gittiggactg aaccatcitig agaataggtg atcaaaagta 900 atttagaatt titatcatca gittiggactg accatcitig accataggitg atcaaaagta 900 atttagaatt titatcacca gittiggactg accatcitig agaataggig atcaaaagta 900 atttagaatt titatcacca gittiggactg accatcitig tittigaggig tittigaaacg accacaggid tatatticac 960 cagsittata gitgaatatig cittigatig tittigaaggig cittigatig tatatticacgig tittigaagacgig tagaataggig tagaataggig accaagaacgid 1020 titaagcaga accacagcaca tatgiaaggig taaacatcitig tattitititi gaataagca 1080 titgagaactgi atgaatcaa gaactggiga titcagaagta tittigaggig cittigatig tittigaagacgic cittigiccaccagagactitig titigaagatgi cittigaggig cittigatig tittigaagaggig cittigatig atcactgagaga tittigaagaa tittigataga accactiat cittigaggig cittigactaa 1320 titigaggig cittigagaagaa agaacagtig tittigaaaaa gittitigaaa accactigagaa tittigaaaa gagaacagtig tittigaaaaa gittitigaaaa gaattitiga gittigaagaaa agatciggig tittigaaaaa gittigaagaa aaattigga atacaagaa ataaaaaaa gittigaagaa aaattigga tittigaaaaa gittigaagaa aaattigga atacaagaa agaatagaga titigaagaa aaattigga atacaagaa agaatagaga titigaagaaga titigaagaa aaattigga atacaagaa titigaagaa aaattigga titigaaaaaa gittigaagaa aaattigaa gitgaatgaaa aaattigaa accittigatig cittigatagaa aaattigaa accittigatig cittigaaaaa agaatagaaa agaacacaaaaaaaaaaaaaaa							
aatcagttta tetttggeag aaatcagtat attataaatg ttaattaat teetaagaat 660 taatggeaat agtgeeteta gitttaggeag gatttaggeag eatttggeag gitttaggeg gatttagge cattetteet eattetgee 720 geecagtaggeggegggegggggggggggggggggggggg							
taatggcaat agtgcctcta gttttagctg agatttagat catctttctt cattctgccc tgttatgcga catttgctat aaatgccaca atgagaattt tttgaaacct gggtcagaga 780 agcctacttt ttatatttaa taagtatgtg attgatgtat tatcattggt ttatgcttat 840 tgcatcttgg ttttatgggg tgttcaaaag aagatagatg agaataggtg atccaaagta 900 atttagaatt ttatcctcca gttgggactg aaccatcttg aactgatgaa tatattcta 960 atttagaatt ttatcctcca gttgggactg aaccatcttg actgatgaa tatattctat 960 atttagaactg agaataggc cgtttcatt atatcggga tttgaggatg atcaaaaggca tatgaagaactg ttagaggaactg acaaccaccacacacagaa tatgtaaggc taaaatttgg ttatttttttt gaataagcaa 1080 tggaaactg atgcattcag gaactggtga ttcagatatc ttctgaaggc cctctccc 1140 tacctatttt gtgaactctg cctgcaatgc cgtcccaagt cagttggtga ctctccaacaccaccaccaccacacacacacacaccacca							-
gettategga cattigotat aaatgocaca atgagaatti titigaaacot gggtcagaga 780 agcotactit titatattaa taagatatga atgatgatat tatcattigs titatgottat 840 tgcatctitg tittataggg tgttcaaaag aagatagatg agaataggtg atccaaagta 900 atttagaatt titatcotcca gttgggactg aaccatcttg aactgatgaa tatatictac 960 caggittata ggtagatatg ccgittcat atatcggtga titigaggtga gacagatagct 1020 tittaagcaga acaacagcca tatgtaagtc taaattiggt tattittett gaataagcaa 1080 tggaaactgt atgcattcag gaactggtga titcagatac titctgaaggc cctctctccc 1140 ctggotccc acagcittig tcctgaatgc cgtcccaagt cgtcccattit ggaatcacg cctgcaatgc cgtcccaagt cctgattit ggaatcaaga tittitagagg tggatcaaa tgctgcttit agcattatit gtgaactct cctgaactc cctagtgtga tittitagagg cctcctcaacacacacacacacacacacacacacacaca							
agoctacttt ttatattaa taagtatgt attgatgta tatcattggt ttatgctat tgatacttgg ttttatgggg tgttcaaaaa aagatagatg agaataggtg atccaaagta 900 caggtttat ggtaggatg accactattg aactgatgaa tattattctac 960 caggtttata ggtagatatg ccgtttcatt atatcggtga tttgaggatg aggcattgag tttttgagg ttgaggatgg atggaggatg tttgaggatg tttgaggatg tttgaggatg ttttgaggatg ttttgaggatgg aggaggatg tttgaggatg tttgaggatg ttttgaggatgg aggaggatg tttgaggatg tttgag							
type technique tentent tente graph of the tente gra							
atttagaatt ttatcctca gttgggactg aaccacttg aactgatga tatattctac ggaagttatat gtgagatatg ccgtttcatt atatcggtga ttttggagtga gacagtagct 1020 tttagagaga acaacagca tatgtaagt tatgaagtga tatgaagatgt tatggaagaga ttttggagtga gacagtagct 1020 tttggaactga atgcattcag gaactggtga ttcagatat ttttgtt tatttttt gaataagcaa 1080 tggaaactgt tggaactct ccgtgaatgc tcctgtgctcc tttggagtgc cctctcccc 1140 tcgtgyctcc acagcttttg tcctgaactg cgtcccaagt cagttggtga ctcctcaacaa 1260 ctcactgaag tctttcagag tggattcaaa tgccgttaca ttttacacaa tagcacttat actttcacaa tagcacttat tcccagtgag ttttaaaaat gcacttaaaa atctgttaca atttttaag gagaggacaga agagtagaa agaattgaa gaattttat gtgagacaga gaattttat aagaacttat tacccatgg gtaatttatg ggaatttaaa caaaaattgga cagtacagaa ataaatggg 1440 aagattgttcg gtaatcaaa agaatttatat actttcacag gtaatttatat acttttactg gtaacacaa agaattgaa acctatgaa gtggattga aaattcaagc cctaactaaa attcatcacg gtagagagaa agattgaaa agattcaaga aaattcaaga caaaaatagga cagtagaggtg 1620 gtggaatga accttggtag ttttaaaaat gcacttaaaa 1500 aaattcacacg gtgagagagaa agattcacga aaattcaaga cagaacacaa agactggagtga ttccctaattaa cattcctcc tccagaagaa aaattcaaga gtgccttttat tactttgatc cctaacacta gccgttatata accttgtgt gtgaacacaca ttacacacta gccctttcca aattacacacg gtcctaattaa cattcaccac gtccagaagaa acctctggt taaagacaaa ttcagagtga ttcctggctac tacacacta gtccagaagaa acctctggt taaagacacac gtcccagagtga attacaggag gccagacaca ttcagaagag ttcccagagtga attacagaga gccagcacaca ttcagaggat tcacacactag gccagcacaca ttcagatgag tcacacacacacacacacacacacacacacacacacaca							
caggittata ggtagatatg ccgttcatt atatcggtga tittggagtga gacagtagct 1020 tittaagcaga acaacagcca tatgitaagtc tatattitgt tattittett gaataagcagaa 1080 tittggaactct atgitaagtc tectgaactc tectgaacaca atgeacttat acaacattgga cagaacaaa atcattacaa 1380 tectgaacaa gaattttata acaacattgga cagaacaaa ataaatggtg 1440 aagaacagaa aggteecett tectccataga acaaaattgga cagaacaaa acaacatgg tectaacaa gaattttata acattcaagc gaattttata acattcaagc gaattttata cattcaagc gaattttata cattcaagc gaattttata cattcaagc acaacatagaa gacttata tatagacatta tectaacaaca gaatttta gagaacaaa aggteecett tectgaacaa gacttetta tatagaacata acaacatga gacttatca acattcaagc gaattttata cattcaaga gaatttta tatagaacatt gacetttata tatagaacatt tectaacaacatg gacttataaa acaacatga gacttataa acattcaaga gacttataa acattcaaga gacttataa acattcaaga gacttata tatagaacata tagaacatta tectgaagaa gactgaagaa acaacacata gacttataaa acattcaaga gactgaacaa acttcaacaca gactgaacaa acttcaagaa gactgaacaa acttcaagaa gactgaacaa acttcaagaa gactgaacaa acttcaagaa gactgaacaa acttcaagaa acaacacata acaacatgaa acctgaacacaa acctgaacacaa acctgaacacaa acctgaacacaa accaacacaa accaacacaa accaacacaa accaacacaa accaacacaaa accaacacaaa accaacacaaa accaacacaaa accaacacaaa accaacacaaa accaacacaaaa accaacacaaaa accaacacaaaa accaacacaaaaaa							
tttaagcaga acaacagcca tatgtaagtc taaatttigt tattitettt gaataagcaa 1080 tggaaactgt atgcattcag gaactggtga ttcagatact ttctgaaggc ctctctccc 1140 ctgtgctcc acagcttttg tcctgaactc ttctgaaggc ctctctccc 11200 tacctatttt gtgaactctg cctgcaatgc cgtcccaagt cagttggtga ctctctaaca 1200 ctcactgaag tctttcaggg tggattcaaa tggtcgtttt agccattaaa atctgttaca 1320 tctgttacat actttcacaa tagcacttaat cctagtggtga ttttaaaaaat gcacttaaaa 1380 attittaaag cagtgcatgc atgcagtaaa acaaattgga cagtacagaa ataaatggtg 1440 aagacagtct cccagttttc ttccccatgg attittat accgtatagaa gatattttg gtgaactaaa aggtcactta cctagttgaga gtgttattcat accgtatagaa gatatttatg gtgaatcaaa gaattttata accgtatagaa gtgcttatc atcttactt 1560 gatattttg gtgaatcaa gaattttat accgtatagaa gtgcttattc atcttactt 1560 gatattttg gtgaacaa aggtccctt tactttgatc cctaacacta gccctttcca aataacactg 1680 ttacttaaaa aggtcccctt tactttgatc cctaacacta gccctttcca aataacactg 1680 tcctaataa cattcctctc tccaagaagta acctctgtta taagcctttg ctgtgcctt 1740 tcttgttgtt tggaacagag tttcgctctt gttgcccagg ctgaagtgca atggacatta 1860 tttgtttgtt tgagacagag tttcgctctt gttgcccagg ctgaagtgca atggacatta 1860 tttttttatc gataccatgt gtctccctgg ttcaagagta tctcctggct cagactcctg gtttcaactt atgagcttat cctggcta atttttgtat ttttagtaga 1980 gatggagttt caccattgttg gccaggctgg tcacgaactc ctgaccatcag gttatccacc 2040 caccttgcc tttcaaacct gtgagctcct ggagaccaa accacaga aggacgcc ttcccgg aagacgaac cttctctccc gcctttccg agaatggac cctggaatgac cctggagaccaaaa tccatggg caccagaaca ccttggaaccaaa cctggaaccaaa cctggaaccaaca tccctgggg caccagaaca cttctctccc ggagaccacaa tccatgggg cctggaaccaaca ccagagaccc ttccccaga gaacacaaca aggacgcct ggaaacaaca ccacagaa agacgcct tccccagaa accacacaga agacgcct tccccagaa ccacacaga agacgcct tcccctggct tcccccggg ccccagaacaaca ccacagaaca accacacaga agacgcct tcccccagaaca ccacacaga agacgccc tccccagaaca ccacacaga agacgccc tccccagaacaca ccacacaga agacgccc tccccagaacaca ccacacaga agacgccc tccccagaaca ccacacaga ccccccagaaca ccacacaca							
tggaaactgt atgcattcag gaactggtga ttcagatact ttctgaagge cctctctece 1140 ctgtgctcc acagcttttg tcctgaacte tcctgtctcc ttgcgatgcc tctgtgccta 1200 ctctactattt gtgaactctg cctgcaatgc cgtccaagt cagttggtga ctcctcaaac 1260 ctcactgaag tctttcaggg tggatcaaa tgctgctttt agccattaaa atctgttaca 1320 ctgttacat actttcacaa tagcacttat cctagtgtggt ttttaaaaat gcacttaaaa 1380 atttttaaag cagtgcatgc atgcagtaaa caaaattgga cagtacagaa ataaatggtg 1440 agaattttg gctgatacaa gaattttat accgtatgaa gtgcttatt tctctttttt tccttccaaa 1500 gatattttg gctgatacaa gaattttat accgtatgaa gtgcttatt tctctttttt tccttcaaa 1500 gatattttg gctgatacaa gaattttat accgtatgaa gtgcttattc atcttcaaaa 1600 gatattttg gctgatacaa gaattttat accgtatgaa gtgcttattc atcttactt 1560 ttacttcagg gtaatttatg agtaatgaaa aaattcaagc attacaaaaa tgtagaggtg 1620 ttacttcatt tcccagaagta accttgtta taagcctttg ctgtgccttg 1740 agattgtct gcttacactt atatgcctat ttaaaagtgt atatttttg ttggacaagg tttcgctctt gttgcccagg ctgaagtgca atggacatat 1860 tttgttgtt tggagacagg tttcgctctt gttgcccagg ctgaagtgca atggacatat 1860 tttgttgtt tggagacagg ttcgcctt gttgcccagg ctgaagtgc atgacactat 1860 tttgttgtt tggagacagg gcaggctgg tcacgacacc ctggaattac aaggaagtc caccatggtg gcaggctgg tcacgaact cttctctcg aaggaggtgc caccatgttg gccaggctgg tcacgaact ctcctgg ttcacagtg ctggacacta 1860 ttttttatc gatacaactg tgaggcttg tcacgaact ctcctgg caccatgtgg caggcacaaa 1980 gatggaggatt caccatgtgt gccaggctgg tcacgaact ctcctctgg attcccttgg attcctctgg attcctctgg attcctctgg gtatcccctg gtaccacaga 2100 ttttttatc gatacaactg tgaggctctct gcctctctcgg aacgtgccc aggaatgcc cttctcccc agaatggc ctcctctccg agaatggc ctcctctccc agaatggc accacagaa accacagaa agctggcacaaaa tccatggtt ctcgtgaacaca accacagaa gatgcgccc accacagaa agctgcccc accacagaa agctggcacaaaa tccatggat ccctctggg ccgggatgg cccggaagggacacaacaca							
tctgtgctcc acagettttg tcctgaacte tcctgtetce ttgcgatgcc tctgtgceta 1200 tacctatttt gtgaactct cctgaatgc cgtcccaagt cagttggtga ctcctcaaaa 1260 ctcactgaag tcttttcgagg tggattcaaa tgtgctttt agccattaaa atctgttacaa 1320 tctgttacaa actttcacaa tagcacttat cctagtgtag ttttaaaaat gcacttaaaa 1380 aagacagtct cccagtttc ttccccatgg atagcactta tcctttttaa tccttctcaaa 1500 gatatttttg gctgatacaa agtaagaaa aatatcagaa atagcacttt tccttcaaa 1500 gatattttg gctgatacaa agtaatgaaa agtaatgaaa agattcagg gtaagagaaa aggtaccctt tccagaagaga acactttg gtgaagagaa aggtacctt tccagaagaga acactttg gtgaagagaa aggtacctt tccagaagaga acactttg gtgaagagaa aggtacttt tccagaagaga acacttgtg tcggatgagggg ttcagagagaa aggtacctt tccagaagaga acacttgtg cggatacaaaa tggaagtgt 1620 gtgaagagaaa aggtacctt tccagaagaga acacttgtg tcggagagtg 1620 gtgaagagaa aggtacctt tccagaagaga acacttgtg cggattattc atctttgt 1740 aggattgtt gggataaaa aggtacctat ttaaaaggtg atattgtgt tggaagagaga acacttgtgaa ggccttttca ataacactg 1680 tttgttgttgtt tggagaagaga acactctgta ttaaaaggtg atttgttgtt tggagaagaga tttcaggagg ttcaagagag ttcaagagga ttcccctgg ttcaagagta acactcggc cagactgag attgtgtgg acacgaagaga acactcggacaaaa tcctctgcc aggcactga 1860 tttgttgttgtt tgagaaagaga tcccctgg ttcaagagta acactcggcc aggagactat 1860 tttttttat gataccattg gccagagagg tcaagagaga catttttgt attttgttg 1800 aggaggaggatt cacacatgg gccacacaa tccccagaagaga caccttggac cgacgacagaa 2100 ttttttaa gataccattg gccaggatga acaggactga acactaggac cactggacc aggacacaa 2100 ttttttaa gataccataa tgagactat ggagacacaa acacgagaa gagagacacaa tccatgaga caccttggac cggagaacaca ctccccagaaga gaagacagcc tttccccaga gagacacacaaaa aggagacacaaa tccatgagt ccccctggaa acccccacaaaa agaaaaaaa ggagacacaaa tccatgagaca acccacaaaa agaaaaaaa ggagacacaaa cacacagaa agacacaca cacacagaa agacacaca cacacagaa agacacacac							
tacctatttt gtgaactctg cctgcaatgc cgtcccaagt cagttggtga ctcctcaaacc 1260 ctcactgaag tctttcaggg tggattcaaa tgctgctttt agccattaaa atctgttaca 1320 atttttaaag cagtgcatga atggacattat cctagtgag ttttaaaaat gcacttaaaa atctgttaca 1320 atttttaaag cagtgcatga atggacatta cctagtgag ttttaaaaat gcacttaaaa 1380 atttttaaag cagtgcatga atggacatta cctagtgag cagtacagaa ataaatggtg 1440 aagacagtct cccagtttc ttccccatgg atagcactt ctcttttta tccttccaaa 1500 gatattttg gctgatacaa gattttta accgtatgaa gtgcttattc atcttactt 1560 ttacttcagg gtaatttatg ggtaatgaaa aaatcaagc gtgcttatc atcttactt 1560 gtgaagcaaa aggtcccct tactttgatc cctaacacta gccctttcca aataacactg 1680 tcctaattaa cattcctcc tactttgatc cctaacacta gccctttcca aataacactg 1680 ttcctaattaa cattcctct tacgctat ttaaaagtg tatttttg attttgtt gagacagag tttcccagg ttcaagatga tctcctggc aggatgacata 1860 ctctgctcac tacaacctg gcctctctg tttaaggtgat tctcctggc aggatgaga atggacatta 1860 ctctgctcac tacaacctg gccagcaca tgcccggct atttttgt ttttgataga 1980 gatggaggtt caccatgttg gccaggctg tcacgaactc ctgacctcag gttatccacc 2040 caccttggcc tttcaaactg tgccgcacca tgcccggcta atttttgt ttttaatcag 1980 gatggaggtt caccatgttg ctggaagtga cacctttggc cggcaagaa atggacacta 1860 ctctttttac gataccatag tgagctctc gcctctcgg aacggactat tttttgataga 1980 gatggaggat caccatggc tttccaagtgg ccacggctg ccggcaagaa atggacacta 1860 ctctttttac gataccaca aggacagac cttctctccc tggacgcctc ctccctggc cggccaagaa 1980 gatggaggag gcagggccc gagagcccct ggagagcacaaa tccatggg cccctttcgg aacgatggc cggcaagaa 2040 ctctctccc gagagcgcct tcccctggc cggcaagaa 2040 ctcccggcaagaa gcagggccct gagagcacaaa tccatgggt ccgggaagcacccc taggaccttg gagagcacaaa tccatgggt cccgggaagcacaaa ccccttttcc 2280 caccgcacca accacagaa gadacacaca accacagaa agccgcctc ctccctgggc ccttgggaagcacaaaa cccctttgcc gaaacacaaa aacagcccc ctccctggcc ctgggaagacaaa 2580 gtgcaattaa cacactgta ccccggaac ccccggaagaccacaaaa accacacaga agccggaccacaaaaa accacacaaa aacaacacaa accacacaaa aacaac							
ctcactgaagtctttcggggtggattcaaatgctgettttagccattaaaatctgtaca1320ctctgttacatactttcacaaattgcacttatcctagtgtagttttaaaaatgcacttaaaa1380atttttaaagcagtgcatgcatgcagtaaacaaaattggacaagtacagaaataaatggg1440aagacagtctcccagttttcttccccatggatagcactttctcttttttatccttctcaaa1500gatattttgtgctgatacaagaattttatacgtatgaagtgcttatcatcttactt1560ttacttcagggtaatttatgagtaatgaaaaaattcaagcattacaaaaatgtagagtgt1620gtgaagcaaaagttccctttcctaacactagccctttccaattacaaaaatgtagagtgt1680tcctaattaacattcctctctccagaagtaacttctgttattaaaagtgatatttttgtctgtgcctt1740agtaggagtttgagacagagtttcgctcttgttgcccaggctgaagtgcaatggacatta1860tttgtttgtttgagacagagtttcgctcttgttgcccaggctgaagtgcaatggcactta1860ctctgtcactacaacctgtgcccggcttttccctggactctccttgcccagcctcctg1920agtaggagtttttcaaactgccaggactgttcacgaactgctccttggccagctttggcagcctctg1920cagttctgctttcaaactgccaggactgtcacgaactgctccttggccagcttggccagccttggcaccttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccagcttggccaggact							
tetgttacat actttcaca tagcacttat cetagtgtag tittaaaaat geacttaaaa attittaaag cagtgcatge atgcagtaaa caaaattgga cagtacagaa ataaatggtg 1440 aagacagtet cecagtitte tececatgg atagcactt cetettitta teettecaaa 1500 gatattittg getgatacaa gaattittat accgtatgaa gtgettate atettatett 1560 titactteagg gtaattitatg agtaatgaaa aaaatecaacta geeettitee atettatett 1560 titactteagg gtaattitatg agtaatgaaa aaatteaage attacaaaaa tgtagagtgt 1620 gigaageaaa aggteeett taettitgate cetaacacta geeettieea aataacactg 1680 teetaattaa catteette teeagaagta acetetgita taageettgt ettigteette 1740 agattytiet getacactt atatgeetat titaaaaggg atatitititititititititititititititititi						,	
attittaaag cagtgcatgc atgcagtaaa caaaattgga cagtacagaa ataaatggtg 1440 aagacagtct cccagttttc ttccccatgg atagcacttt ctctttttt tcctcaaa 1500 gatattttgt gctgatacaa gaattttat accgtatgaa gtgcttattc atcttcaatgt 1560 ttacttcaagg gtaatttatg agtaatgaaa aaattcaagc attacaaaaa tgtagagtgt 1620 gtgaagcaaa aggtccctt tactttgatc cctaacacta gccctttcca aataacactg 1680 tcctaattaa cattcctctc tccagaagta acctctgtta taagccttgt ctgtgccttc 1740 agattgttct gcttacactt atatgcctat ttaaaaatgg atatttttgt attttgttg 1800 ctctgttatt tgagacagag tttcgctctt gttgcccagg ctgagtgca atggcactat 1860 ctctgttcac tacaacctg gtctccctgg ttcaaagtgat tctcctggct cagcctcctg 1920 agtaggttg caccatgttg gccagccaca tgcccggcta atttttgat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacgaactc ctgacctcag gttatccacc 2040 caccttggcc tttcaaactg ttgggattac aggcactct tggactcag accetttgcccaggtcagatcaccc aggcagagac cttctctccc tggacgccc ctggccaagaa cccctttgcccaggtcagagacgagac	_				_	_	
aagacagtct cccagtttic ttcccatgg atagcacttt ctcttttta tccttccaaa 1500 gatattttgt gctgatacaa gatttttat accgtatgaa gtgcttattc atcttactt 1560 ttacttcagg gtaatttatg ggtaatgaaa aaattcaagc attacaaaaa tgtaagggtg 1620 tcctaattaa cattcctctc tccagaagta acctctgtta taagccttgt ctgtgccttc 1740 agattgttct gcttacactt atatgcctat ttaaaagtgt acctttgta taagccttgt ctgtgcttc 1740 agattgttct gcttacactt atatgcctat ttaaaagtgt atattttgt attttgtttg 1800 tttgtttgtt tgagacagag ttttcgctctt gttgcccagg ctgaagtgca atggcactat 1860 ctctgtgcttc taccactggt tcaagtgat tctcctggcc cagccactat 1860 ctctgtgctt taccagtggt tcacagtgat tctcctggcc cagcctcctg gttcaagtgat tctcctggc cagcaccac tgcccaggat atttttgat tttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacaggatg cacgcaccac tgcccggcaa atttttgat tttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacggaactc ctgacctcag gttaccacc 2040 caccttggcc tttcaaaccg ctgggattac aaggcatgag ccactgtgcc cggccaagaa 2100 tttttttac gatacactag tgagctctct gcctctcgg acgactggc ccggccaagaa 2100 tttttttac gatacacca agaatggcc cttctctccc tggacgcct tcccctggtc tggaatcttc 2220 cagttctgcc caggggctg gagaccaaaa tccatggatg ctcccttgtg accacgggct tccccggct accacgacga gcagggccc gagaaaaaggc aaaacgccc ctccctggtc tcggaatctc 2280 tggaggccctg accacgaaga gcagggccc gagaacaaaa tccatggata ccccttttcc 2280 tggagccctga accaccacaa agaatggca aaaacgccc ctctcctggc ccttctggc accacgatgg cccccggcaacaa accaccacaaa agacgccc tccccggaa agaccaccac accacacaaa agaccgccc ctccctggcc ccttgggatgg 2520 cccaaaaaa ccacacgtata cacacacaaa agaccgccc tcccctggcc ccttggaaccc 2640 tccaggaaacc accaccacaaaa agaccccc tccccggaa agggacctc caggcagcc 2700 ccaaaaaaaaaaa ccacacaaaaa acccccaacaaaa agacccccc aaacacaaaa accaccaaaaaaaa							
gatattitgt gctgatacaa gaatttitat accgtatgaa gtgcttattc atcttatctt tacttcatcy gtaatttatg agtaatgaaa aaattcaagc attacaaaaa tgtagagtgt 1620 gtgaagcaaa aggtcccctt tactttgatc cctaacacta gccctttcca aataacactg 1680 tcctaattaa cattcctct tccagaagta acctctgtta taagccttgt ctgtgccttc 1740 agattgttct gcttacactt atatgcctat ttaaaagtg atatttttgt attttgttg 1800 tttgtttgtt tgagacagag tttcgctct gttgcccagg ctgaagtgca atggcactat 1860 ctctgctcac tacaacctg gtctccctgg ttcaagtgat tcccctggt cagcctcctg 1920 agtagctagg attacaggtg cacgccacac tgcceggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacggactc ctgcaggctg gtatccacc 2040 caccttggcc ttcaaactg tgagctctct ggagctcg accgtggatga accgagactc ctgagactcc agactggcc cggccaagaa 2100 ttttttatc gataccatag tgagctctct tcccctggt tcacgatgtc cagccacac agagagact cttcctccc tggacgcctc tcccctggtc tggaatttc 2220 cagttctgc agaattggcc ctggaggccg gagacaaaa tccatgagtt ccacggatgc ctgagggtggag agaggcacaaa tccatgagtt cccatggtg tggaatcttc 2220 cagttctgc accgcacac gcagacacac accacaga agacgcct ctcccctgga ccctttttc 2280 tggagcccta ctcccacca gaattgtcag tggccttca cccacgatggt ctcccgtc ctcccctgga cccttttcc 2460 acccgcacta ctcctcacca gaattgtcag tggccttca cccacgatgg cccttggagtgg 2520 tcaccgaaaca caccacaga agacgcct ctcccctggac cccggaaccc ctccacgaaccc accaccagaa accaccacaa agtcgaaccc accacgagac ccccggaaccc accaccaga agtcgaaaccc accaccaga agtcgaaaccc accaccaga agtcgaaaccc accaccaga agtcgaaaccc accaccaga agtcgaaaccc accaccaga agtcgaaaccc accaccacaa agtcgaaccc accaccacaa agtcgaaccc accaccacaa agtcgaaaccc accaccacaa agtcgaaccc accaccacaa agtcgaaccc accaccacaa agtcgaaccc accaccacaa agtcgaaccc accaccacaa accaccacaacccaccaccacca	_				-		
ttacttcagg gtaatttatg agtaatgaaa aaattcaagc attacaaaaa tgtagagtgt gtgaagcaaa aggtccctt tactttgatc cctaacacta gccctttcca aataacactg 1680 tcctaattaa cattcctctc tccagaagta acctctgtta taagccttgt ctgtgcctc 1740 agattgttct gcttacactt atatgcctat ttaaaaagtg atatttttgt atttgttgt ttgagacagag tttcgctctt gttgcccagg ctgaagtgca atggcactat 1860 ctctgctcac tacaacctgt gtctccctgg ttcaagtgat tctcctggct cagcctcctg 1920 agtagcagg attacaggtg cacgcacca tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcaaggac ccactggcc cggccaagaa 1980 gatggagttt caccatgttg gccaggctgg tcacgaactc ctggactca ggaccacca tgcccggcta atttttgtat ttttagtaga 1980 caccttggcc ttcaaactg tggagctcct gccctcttgg accgacacca ggaccacaaa tccactggcc ttcacaactg tggagctcct gccctcttgg accgacacaa caccttgcc aggaccacaaa tccactggcc ttccccaga gttatccacc 2040 caccttggcc ctggagctcc ttccccaga gcccctcttgc accgacacaa ccctttgcta 2220 cagttctgc agaattggcc ttcccagaa gccccacaaaa tccatgagtt ctcacgttgaa cccctttttc 2280 gaaggtgaga gcagggccc gagaacaaaa tccatgagtt ctgtgtaaccc tagacctttg 2340 gaaggtgaga gcagggcccc accacacaaa agaccaccaa agaccaccaaaa tccacagaaccaaa agaccaccaaaa cccccttggag ccccggaaccaccaaaaaccacaaaaaccacaaaaccaaaaaccaaaa							
gtgaagcaaa aggtccctt tactttgatc cctaacacta gccctttcca aatacactg 1680 tcctaattaa cattcctctc tecagaagta acctctgtta taagccttgt ctgtgccttc 1740 agattgttct gcttacactt atatgcctat ttaaaagtgt atatttttgt atttgtttg 1800 tctgtgctccac tacaacctgt gtctccctgg ttcaagtgat tctccttggct cagccaccac tggccggcac atttttgttg ttgagacagag ttcacctgg ttcaagtgat tctccttggct cagccaccac tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcaccgaccac tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcaccgaccac ctgacctcag gttatccacc 2040 caccttggcc tttcaactg ctgggattac aaggcattgag ccacttggcc cggccaagaa 2100 ttttttatac gataccatag tgagctctct ggacgcctc ttggaacctc tctctccc tggacgcctc cagactctcgg aacgatgagtctcgg gagaccaaaa tccatgggc ctgggggcgc agaacgggggggggg							
tcctaattaa cattcctct tccagaagta acctctgtta taagccttgt ctgtgccttc gettacactt atatgcctat ttaaaagtgt atatttttgt attttgtttg 1800 tttgtttgtt tgagacagag ttttcgctct gttgcccagg ctgaagtgca atggcactat 1860 ctctgctcac tacaacctgt gtctccctgg ttcaagtgat tctcctggct cagcctcctg 1920 agtagctagg attacaggtg cacgcaccac tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacgaactc ctgacctctag gttatccacc 2040 caccttggcc ttcaaactg tgggctctct gccttcgg accgttggc cggccaagaa 2100 tttttttatc gataccatag tgagctctct gccttctcgg accgttgtcc cggccaagaa 2100 tttttttatc gataccatag tgagctctct tcccagat tccccttggtc tggaatcttc 2220 cagttctgc agaattggcc tttcccagat gctgcaaact tccatgagt cggaatcttc 2220 tggaggggag gcaggggcct gagaaaggc aaaacgcct ctctcctcga tggacccttt 22400 acccgcccta ctcctcacca gaattgtcag tggccttca ctctccctga tgatacccct 2400 acccgcccta ctcctcacca gaattgtcag tggccttca ccctctgggc cctgggatgg 2520 tcaccgaaga gaagcacgct gcccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactc caggacttgt ccttgatcct 2640 tccaagaaacc accacgat tccaggac ctcaggac ctcaagacaccacacacacacacacacacacacacacaca							
agattgttct gcttacactt atatgcctat ttaaaagtgt atatttttgt attttgtttg ttggtacactat tgaagaagtgat tcccaggatgatat tccctggct cagcctctg ttcaagtgat tccctggct cagcctctg 1920 agtagctagg attacaggtg cacgcacca tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tccacgactc ctgacctcag gttatccacc 2040 caccttggcc ttcaaactg ctgggattac aaggcatgac cactttggcc cttcaaactg tgagctctct gcctcttcgg aacgatgcc accttggcc aggacacaa tcctctccc tggacctcag gttatccacc 2040 caccttggcc aagaactgagact cttctcccc tggacgtgg accatttggcc caggcagaac 2100 ttttttatc gataccatag tgagctctc tggacgcct tcccctggtc tggaatcttc 2220 cagttctgcc aggattggcc ctgggggtg gagacaaaa tccatgggt ctggaacct tccagttgaa cccctttttc 2280 tgtgtggcc ctggggctg gagacaaaa tccatgagt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggcct gagaacagaa agcctctcaggt tctccctga tggataccct 2400 acccgccta ctcctcacca gaattgtcag tggcctttca cacacagtggt ccttcctgcc 2460 tgagccctgc actgtccag accacagaa agcctggtca cctctggggg cctggaagg gaagacagac accacacaga agtctggtca cctctggggg cctgggatgg 2520 tcaccgaaaa cacacgtta ctgccgaaac tccatgaaat cacactgta ctgccagaac cccccggag agggacctc caggacgcc 2640 tccagaaacc accacgtca caccttggag cccccggag agggacctc caggacgcc 2700 ctcaaagaac tccatgaaat caccctggg gaggtggga cagggacctc caggacatac 2820 cggccacaca accaccaca accttgtc ccccaggagac cagggaccat ctctggaac 2940 tacagaacga gaaaaaaaa aaaaaaaaa agggcggcgc tctagaggat ccctcgaggg 3000 gccccagctt acccagctt acccccagaggct cctccagagg tcctcccagaggcc ctcaagagac ccccaggac ccccaggagc cctccagagg ccccagagac cccccagagg ccccagagac cccccagagg ccccagagac ccccagagac ccccagagac ccccagagac ccccagagac cccccagagac ccccccagagac cccccagagac cccccagagac cccccagagac ccccca							
tttgtttgtt tgagacagag tttcgctctt gttgcccagg ctgaagtgca atggcactat 1860 ctctgctcac tacaacctgt gtctccctgg ttcaagtgat tctcctggct cagcctcctg 1920 agtagctagg attacaggtg cacgccacca tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacgaactc ctgacctcag gttatccacc 2040 caccttggcc ttcaaacca ctgggattac aaggcatgag ccactgtgcc cggccaagaa 2100 ttttttatc gataccatag tgagctctct gcctcttcgg acacgatgtcc actttgctc agaactctc tggacgccc cagttatect 2220 cagttctgcc agaattaggcc ttcccccag tggacgccc tggacgccc tggaggcgcc tggagacaaaa tccatgagt ctgtgtaccc tagaccttt 2220 cagttctgg gcagggcgcc gagacaaaaa tccatgagt ctgtgtaccc tagaccttt 2230 gaaggtgaga gcagggccct gagaacaaaa tccatgagt ctgtgtaccc tagaccttt 22400 acccgccta ctcctcacca gaattgtcag tggccttca ccacagtggt ccttcctggc 2460 tgagccctgc actgtccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaagac accactgtta ctgccagacc ctgaaactc ctgaaactc caggacttgt ccttgatcct 2640 tccagaaacca accactgtta ctgccagaac cccccggga ggaggaccaca accacaga gtagaactgc ttgatgaaccc caggaccttg ccttgaaccc cagcaggcc 2760 ctcaaagaac cacaccaca acctctgtct ccccccaggaggaccat ctctctccc 2880 tccacccagca gaaaaaaaaa acccccaccaga gagggggccgc tctagaggac cctctgggg cagggaccat ctctcccc 2880 tccaccccca ctccacca gaaaaaaaaa agagcgggccgc tctagaggat ccctcgaggg caggaccacct ctccagagg caggaccacc ctccagagg caggaccacc ctccagagg caggaccaccc ctccagaggc ctctcagagg caggaccaccc ctccagagg caggaccacc ctccagaggc ctccagagac ccccccagagacc ctccaccccagacc gctcccagg gaaaaaaaaa acccaccaccaccacccccccaaccaccac							
ctctgctcac tacaacctgt gtctccctgg ttcaagtgat tctcctggct cagcctcctg 1920 agtagctagg attacaggtg cacgccacca tgcccggcta atttttgtat ttttagtaga 1980 gatggagttt caccatgttg gccaggctgg tcacgaactc ctgacctcag gttatccacc 2040 caccttggcc tttcaaactg ctgggattac aaggcatgag ccactgtgcc cggccaagaa 2100 tttttttatc gataccatag tgagctctct gcctcttcgg aacgatgtcc actttgcta 2160 tgatcaaccc aagcaggact cttctctccc tggacgcctc tccctggtc tggaatcttc 2220 cagttctgcc agaattggcc cttggggctgc gagaccaaaa tccatggtt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggccct gagaaaaggc aaaacgcctc ctctcctga tgataccct 2400 accegcccta ctcctcacca gaattgtcag tggcctttca ccacagtggc cctgggatgg 2520 tcaccgaaga gaagcagct gtccccaga agtctggtca cctggggcg cctgggatgg 2520 tcaccgaaga ccacagtaga cccctgggc ctggaaacca 2580 gtgcaattaa cacactgtta ctgacactac caggaacttc ctggacgccc caggcagacc 2700 ctcaaagaac cacaccaca acccccgggg tggacactac cacccgggg tcacaggac aacccccac acccccgagg caggaacctc cccccgagg tcacaggacc aacccccac acccccggg tgggcccgc tgggagaccac cccccgggg tcccccgga caggaacctc cccccgaggc caggaccac 2760 gatgaagccc aacccccac acccttgtct ccccccagggc tcgggatgtc tccagcagcc 2820 cggccaccca gcttcccag tgggcccgg aggggccgc tctagaggat ccctcgaggg 3000 gccccagctt acccgtcat accccccca accccccc accccccc ctcatagtgg tcctccagggg 3000 gccccagctt accccgtcat acccccccacca accctcccc ctatagtgag tcgtattata 3060							
agtagctagg attacaggtg cacgccacca tgcccggcta attititgtat tittagtaga 1980 gatggagtit caccatgitg gccaggctgg tcacgaactc ctgacctcag gttatccacc 2040 caccitiggec tittcaaactg ctgggattac aaggcatgag ccactgigec cggccaagaa 2100 tittitatic gataccatag tgagctetet gcctcticgg aacgatgtcc actitigetia 2160 tgatcaaccc aagcaggact citetecec tggacgcctc tcccctggtc tggaatctic 2220 cagiticigec agaatiggec tittcccagat gctgcaaact tccagitigaa cccctititic 2280 tgtgtgggec ctgggggtgc gagaccaaaa tccatgagit ctgtgtaccc tagacctitig 2340 gaaggtgaga gcagggecct gagaaaaggc aaaacgcctc ctctccctga tgatacccti 2400 acccgccta ctcctcacca gaatigicag tggcctitica ccacagiggg cctgggatgg 2520 tcacgaaga gaagcacgci gtcccggct tecetgget tecetgggc cctgggatgg 2520 tccacgaaga gaagcacgci gccccgtci tcgcagaacc caggactici ctggcaaaac 2580 gtgcaattaa cacactgita ctgccgaagc cccccggag agggacctcc caggccgagcc 2700 cccaagaaacc cacaccaa accitigici ccccagggg tcggaatgic ccacggagc 2820 cggccacgca gcttcccagg tgggctcgg gaggtgggag cagggaccat ctctggagg 2820 cggccacgca gcttcccagg tgggctcgg gaggtgggag cagggaccat ctctggagg 2940 tacaggacgt accacgca gaaaaaaaa aaaaaaaaa ggggcggccg ctatagggg tcgtattata 3060							
gatggagttt caccatgttg gccaggctgg tcacgaactc ctgacctcag gttatccacc 2040 caccttggcc tttcaaactg ctgggattac aaggcatgag ccactgtgcc cggccaagaa 2100 ttttttatc gataccatag tgagctctc gcctcttcgg aacgatgtcc actttgctta 2160 tgatcaaccc aagcaggact cttctctcc tggacgcctc tcccctggtc tggaatcttc 2220 cagttctgcc agaattggcc tttcccagat gctgcaaact tccaggttgaa cccctttttc 2280 tgtgtggccc ctggggctgc gagaccaaaa tccatgagtt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggccct gagaaaaggc aaaacgcctc ctctccctga tgatacccct 2400 acccgcccta ctcctcacca gaattgtcag tggcctttca ccacagtggt ccttcctgcc 2460 tgagccctgc actgtcccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accacggtcg gcacttggag cccccgggg agggacctcc caggcagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc accccagg tgggctcggg gaggtgggag cagggaccat ctctggcgg 2820 cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accctcccc agccagatac ggaataaaac 2940 tacaggacgca gaaaaaaaa aaaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtcat acgcgtcat agcccccc ctatagtgag tcgtattata 3060							
caccttggcc tttcaaactg ctgggattac aaggcatgag ccactgtgcc cggccaagaa 2100 ttttttatc gataccatag tgagctctct gcctcttcgg aacgatgtcc actttgctta 2160 tgatcaaccc aagcaggact cttctccc tggacgcctc tcccctggtc tggaatcttc 2220 cagttctgcc agaattggcc tttcccagat gctgcaaact tccagttgaa cccctttttc 2280 tgtgtggccc ctggggctgc gagaccaaaa tccatgagtt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggcct gagaaaaggc aaaacgcctc ctctcctga tgataccct 2400 acccgcccta ctcctcacca gaattgtcag tggcctttca ccacagtggt ccttcctgcc 2460 tgagccctgc actgtccaga accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accacggtcg gcacttggag cccccggag agggacctcc cagccgagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aacaccaca acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatcac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcggagg 3000 gccccagctt acgcgtgcat gcgacgtcat agcctctcc ctatagtgag tcgtattata 3060							
tttttttatc gataccatag tgagctetet geetettegg aacgatgtee actttgetta 2160 tgateaacce aageaggaet ettetetee tggaegeete teecetggte tggaatette 2220 cagttetgee agaattggee ttteecagat getgeaaact teeagttgaa eeeettte 2280 tgtgtggee ettggggeete ettggggeete etggageete tggaatette 2280 tgtgtggee etggggeete etggggeete etggggeete 2340 gaaggtgaga geagggeete gagaaaagge aaaaegeete eteteetgg tgatacceet 2400 acceegeeta eteeteea gaattgteag tggeetttea eeacagtggt eetteetgee 2460 tgageeetge actgteecag accacacaga agtetggtea eetegggegegeete etgggatgg 2520 teacegaaga gaageaeget gteecegtet teteetgget tetegeagaa aategaacaa 2580 gtgeaattaa eacaetgtta etgeegaage etgaaaetee eaggaeettgt eettgateet 2640 teeagaaace accaeggteeg geaettggag eeeeeggag agggaeetee eaggaeetee 2700 eteaaagaac teeatgaaat eaggaaetge ttgatgaaat gtateteett gtacetggaa 2760 gatgaageee aacaeceae acctetgtet eeeeeaggge tegggatgte teeageagee 2820 eggeeaegea getteeeagg tgggeteggg gaggtgggag eagggaeeat etetgteeee 2880 teeaegeee gaaaaaaaa aaaaaaaaa gggeggeege tetagaggat eeettgggggg 3000 geeeeagett aeggeteat gegaeteat ageteteee etatagtgag tegtatataa 3060	gatggagttt	caccatgttg	gccaggctgg	tcacgaactc	ctgacctcag	gttatccacc	
tgatcaaccc aagcaggact cttctctcc tggacgctc tccctggtc tggaatcttc 2220 cagttctgcc agaattggcc tttcccagat gctgcaaact tccagttgaa cccctttttc 2280 tgtgtggccc ctggggctgc gagaccaaaa tccatgagtt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggccct gagaaaaggc aaaacgcctc ctctccctga tgataccct 2400 acccgcccta ctcctcacca gaattgtcag tggcctttca ccacagtggt ccttcctgcc 2460 tgagccctgc actgtcccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accacggtcc gcacttggag cccccggag agggacctcc caggccgagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aacaccaca acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttccaagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcggagg 3000 gccccagctt acgcgtgcat gcgacgtcat agcctctcc ctatagtgag tcgtattata 3060							
cagttctgccagaattggcctttcccagatgetgcaaacttecagttgaacecetttte2280tgtgtggcccctggggctgcgagaccaaaatecatgagttctgtgtaccctagacetttg2340gaaggtgagagcagggccctgagaaaaggcaaaacgcctcctctccctgatgataccct2400acccgccctactcctcaccagaattgtcagtggcctttcaccacagtggtccttcctgcc2460tgagccctgcactgtcccagaccacacagaagtctggtcacctctgggcgcctgggatgg2520tcaccgaagagaagcacgctgtccccgtcttctcctggcttctgccagaaaatcgaacaa2580gtgcaattaacacactgttactgccgaagcctgaaactcccaggacttgtccttgatcct2640tccagaaaccaccaggtccggcacttggagcccccggagagggacctcccagccgagcc2700ctcaaagaacctccatgaaatcaggaactgcttgatgaaatgtatctccttgtacctggaa2760gatgaagcccaaacacccacacctctgtctcccccagggctcgggatgtctccagcagcc2820cggccacgcagcttcccaggtgggctcggggaggtgggagcagggaccatctctgtccc2880tccaccctcactccatccacctcggagaccaccetccccagccagatacggaataaaac2940tacagacgcagaaaaaaaaaaaaaaaaaaaagcccccccctaagggaccctcggaggcctcggagggccccagcttacgcgtgcatacgcagtcatacctctgagggcctcagaggaccctcggagg							
tgtgtggccc ctggggctgc gagaccaaaa tccatgagtt ctgtgtaccc tagacctttg 2340 gaaggtgaga gcagggcct gagaaaaggc aaaacgcctc ctctcctga tgataccct 2400 acccgcccta ctcctcacca gaattgtcag tggcctttca ccacagtggt ccttcctgcc 2460 tgagccctgc actgtcccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accacggtcc gcacttggag cccccggag agggacctcc caggccgagc 2700 ctcaaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aaacaccaca acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accctcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060							
gaaggtgaga gcagggcct gagaaaaggc aaaacgcctc ctctcctga tgataccct 2400 acccgcccta ctcctcacca gaattgtcag tggcctttca ccacagtggt ccttcctgcc 2460 tgagccctgc actgtcccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accaggtccg gcacttggag cccccggag agggacctcc cagccgagcc 2700 ctcaaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aaacaccac acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060	cagttctgcc	agaattggcc	tttcccagat	gctgcaaact	tccagttgaa	ccccttttc	2280
accegeceta etecteaca gaattgteag tggeetttea ceaeagtggt eetteetegee 2460 tgageeetge actgteecag accacacaga agtetggtea eetetggeg eetgggatgg 2520 teaecgaaga gaageaeget gteecegtet teteetgget tetgeeagaa aategaacaa 2580 gtgeaattaa eacaetgtta etgeegaage etgaaactee eaggaettgt eettgateet 2640 teeaagaace accaggteeg geaettggag eeceeggag agggaeetee eageegagee 2700 eteaaagaac teeatgaaat eaggaactge ttgatgaaat gtateteett gtaeetggaa 2760 gatgaageee aacaeceae acctetgtet eeceeaggge tegggatgte teeageagee 2820 eggeeaegea getteeeagg tgggeteggg gaggtgggag eagggaeeat etetgteeee 2880 teeaeeetea eteeateeae eteeggaaee acceteeeee ageeagatae ggaataaaae 2940 taeagaegea gaaaaaaaa aaaaaaaaa gggeggeege tetagaggat eegtatataa 3060							
tgagccctgc actgtccag accacacaga agtctggtca cctctgggcg cctgggatgg 2520 tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accaggtccg gcacttggag ccccccggag agggacctcc cagccgagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aaacaccac acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060							2400
tcaccgaaga gaagcacgct gtccccgtct tctcctggct tctgccagaa aatcgaacaa 2580 gtgcaattaa cacactgtta ctgccgaagc ctgaaactcc caggacttgt ccttgatcct 2640 tccagaaacc accaggtccg gcacttggag ccccccggag agggacctcc cagccgagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aaacaccac acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060	acccgcccta	ctcctcacca	gaattgtcag	tggcctttca	ccacagtggt	ccttcctgcc	
gtgcaattaa cacactgtta ctgccgaage ctgaaactce caggacttgt ccttgatcct 2640 tecagaaace accaggtceg gcacttggag ceeeceggag agggacetee cageegagee 2700 ctcaaagaac tecatgaaat caggaactge ttgatgaaat gtateteett gtacetggaa 2760 gatgaageee aaacacecae acctetgtet eeeceaggge tegggatgte tecageagee 2820 eggecacgea getteecagg tgggeteggg gaggtgggag cagggaceat etetgteeee 2880 tecacectea etecatecae etecggagaee acceteeeee ageeagatae ggaataaaae 2940 tacagaegee gaaaaaaaaa aaaaaaaaaa gggeggeege tetagaggat eeetetgaggg 3000 geeceagett acgegtgeat gegacgteat agetetetee etatagtgag tegtattata 3060	tgagccctgc	actgtcccag	accacacaga	agtctggtca	cctctgggcg	cctgggatgg	
tccagaaacc accaggtccg gcacttggag ccccccggag agggacctcc cagccgagcc 2700 ctcaaagaac tccatgaaat caggaactgc ttgatgaaat gtatctcctt gtacctggaa 2760 gatgaagccc aaacaccac acctctgtct cccccagggc tcgggatgtc tccagcagcc 2820 cggccacgca gcttccagg tgggctcggg gaggtgggag cagggaccat ctctgtccc 2880 tccaccctca ctccatccac ctcggagacc accetcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060	tcaccgaaga	gaagcacgct	gtccccgtct	tctcctggct	tctgccagaa	aatcgaacaa	2580
ctcaaagaactccatgaaatcaggaactgcttgatgaaatgtatctccttgtacctggaagatgaagcccaaacacccacacctctgtctccccagggctcgggatgtctccagcagcccggccacgcagcttcccaggtgggctcggggaggtgggagcagggaccatctctgtccctccaccctcactccatccacctcggagaccaccetccccagccagatacggaataaaaatacagacgcagaaaaaaaaaaaaaaaaaaagggcggccgctctagaggatccctcgaggggccccagcttacgcgtgcatgcgacgtcatagctctctcctatagtgagtcgtattata	gtgcaattaa	cacactgtta	ctgccgaagc	ctgaaactcc	caggacttgt	ccttgatcct	2640
gatgaagcccaaacacccacacctctgtctcccccagggctcgggatgtctccagcagc2820cggccacgcagcttcccaggtgggctcggggaggtgggagcagggaccatctctgtccc2880tccaccctcactccatccacctcggagaccaccetccccagccagatacggaataaaaa2940tacagacgcagaaaaaaaaaaaaaaaaaaagggcggccgctctagaggatccctcgaggg3000gccccagcttacgcgtgcatgcgacgtcatagctctctccctatagtgagtcgtattata3060							2700
cggccacgca gcttcccagg tgggctcggg gaggtgggag cagggaccat ctctgtcccc 2880 tccaccctca ctccatccac ctcggagacc accctcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaaa aaaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctctcc ctatagtgag tcgtattata 3060	ctcaaagaac	tccatgaaat	caggaactgc	ttgatgaaat	gtatctcctt	gtacctggaa	2760
tccaccctca ctccatccac ctcggagacc accctcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060	gatgaagccc	aaacacccac	acctctgtct	ccccagggc	tcgggatgtc	tccagcagcc	2820
tccaccctca ctccatccac ctcggagacc accctcccc agccagatac ggaataaaac 2940 tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctccc ctatagtgag tcgtattata 3060	cggccacgca	gcttcccagg	tgggctcggg	gaggtgggag	cagggaccat	ctctgtcccc	2880
tacagacgca gaaaaaaaa aaaaaaaaa gggcggccgc tctagaggat ccctcgaggg 3000 gccccagctt acgcgtgcat gcgacgtcat agctctctcc ctatagtgag tcgtattata 3060							2940
gccccagctt acgcgtgcat gcgacgtcat agctctctcc ctatagtgag tcgtattata 3060							3000
							3060
			-				3070

<210> 30 <211> 2227 <212> DNA

<213> Homo sapiens

<220>

```
<221> SITE
  <222> (289)
  <223> n equals a,t,g, or c
  <400> 30
  cggacgcgtg ggtcgaccca cgcgtccggg aaaaarggaa aaratgccgt gtaaaatctc
                                                                         60
 gttctgtgtc tgaattgccg taggctcaga tcttcatttg aggttctgtg tctgaattgc
                                                                        120
 cgtaggetca gatetteatt tgaggttatg ttetataagt taaegttgat ettgtgtgag
 ctttcggtag ctggagtaac acaggeggcc tcacagcgac ctctccagcg ccttccaagg
                                                                        180
                                                                        240
 cacatetgea gecagegtaa teeteetggg agatgeetee teaaggeent getecagace
                                                                        300
 acgtggggar ggcctgacar ccaattccca ggctgtcccc acccttgrag agtgacccta
 aacgctagac agatggggaa tgggaaagaa aagaaagctg cagacctcaa gttaaaattc
                                                                        360
                                                                        420
 cctcaaaaac gtttttattt atctgctttt tctgaaagga taaaggcttt ttgaaaatta
                                                                        480
 ttttctaaca aataacatga acacttctag aaaccctaga aaaacacaaa gtattcaaaa
 tagaaagaaa aattacccat tactctttaa gccagcatta tccattgcgg tgcttttgga
                                                                        540
                                                                        600
 gttgggtgag gccgtagcct ctgccaagtc aaggagcccg gtggtggctg tggcattcct
 gcagggttgt tttttttct ttgagatgga gtctcactct tgtcacccca gctggaatgt
                                                                        660
                                                                        720
 ggtggtgtaa acageteact gcageettga ecetgagget caagegatee ttetgeettg
                                                                        780
 gcctcctgag tagctgggat cccaggcgag agtcaccaca ccctgtccat gttcctgcag
 gtettgatat gegaggaege tgtgtettee etgecacatt ttettettet ttettgagae
                                                                        840
 agaccettge tecateacce aggecagagt gtggtsgtge gaacaegget caetgeagee
                                                                       900
                                                                       960
 togaccetca ggetcaageg atcetcaege eteggaceee caaagtgetg ggatcaeagg
                                                                      1020
 cgagagtcac catgetggcc tgaatettca gggtatttta cggttgaagt gtcacttact
 tarccatscc tgtttcaaga gtgtaggtgg tcaccctgtc tctgycgctg acctggcctg
                                                                      1080
 gaccctcggc tgtgagaggg aggggtgggc tgggctggag gaacctraag ccctcgtgat
                                                                      1140
 gtcacaagcc catctggctg ggcatcccct gctgtgtcct gagctgcaca tgccccaggt
                                                                      1200
                                                                      1260
 ggccccaca gcagaggcga gccactgrag ggtgragggc ttccacggac ggtcttcagg
 ggragaagaa gggcccaggc ccccaggaga ctcaggagac cagagcctgg ggtcaggggc
                                                                      1320
                                                                      1380
 tyagcagggg ctyarccagg gctggatgtc cggagccagc cccgmagecc tgkgktcttt
                                                                      1440
 gttcttcgca ctcccaccgt ccgtgtgaac agetccagcc ccacctgcgc ctccctgtgc
                                                                      1500
 tgggctccat cagggagccc agaagacgtg tgtgcttctg aaattgggtc cctacatgcc
                                                                      1560
 tttgtcccag tgcaccttgc tccttccatt tactatcgag atttaaatgc ctgtttctc
cccagaggtt gacggatata ttcagacgtt acgacacgga tcaggacggc tggattcagg
                                                                      1620
tgtcgtacga acagtacctg tccatggtct tcagtatcgt atgaccctgg cctctcgtga
                                                                      1680
agagcagcac aacatggaaa gagccaaaat gtcacagttc ctatctgtga gggaatggag
                                                                      1740
                                                                      1800
cacaggtgca gttagatgct gttcttcctt tagattttgt cacgtgggga cccagctgta
catatgtgga taagctgatt aatggttttg caactgtaat agtagctgta tcgttctaat
                                                                      1860
                                                                      1920
gcagacattg gatttggtga ctgtctcatt gtgccatgag gtaaatgtaa tgtttcaggc
attetgettg caaaaaate tateatgtge ttttetagat gtetetggyt etatagtgea
                                                                      1980
                                                                     2040
aatgctttta ttagccaata ggaattttaa aataacatgg aacttacaca aaaggctttt
                                                                     2100
catgtgcctt acttttttaa aaaggagttt attgtattca ttggaatatg tgacgtaagc
                                                                     2160
2220
aaaaaaa
                                                                     2227
<210> 31
<211> 1288
<212> DNA
<213> Homo sapiens
<400> 31
ccccgggct gcaggaattc ggcacgagcc tgacctcccc agcctcatct ctcctcctc
tgctctcgcc ctctgtgctc cagccaacgt ggcctgtcac tcgtccacct gccatactgt
                                                                       60
cetgacteca ggcetttgcc tgtgctatgg cetetgttgg gaccactett gtetetecee
                                                                      120
                                                                      180
tgctgtgtct gctaattccc actcgtgtca gtgatccatg gctgcagaac acaccactcc
                                                                      240
atccatggaa aacaatcaca atcattgatt actatctctc cctgggcttc ctggggtgga
                                                                      300
ctgggctcag ctgggtggtt cactttgggg cctcagcagt catgggcaga cagtggctgg
                                                                      360
```

```
ggtcactgca aagacttccc tgcatctctg gcagttgatg ctggctgtca tctgagacac
                                                                         420
 ctacccaggg cetetecetg gggcetggge teetgettag ettggttggg aggetecaag
                                                                         480
 accaacatcc caagaaagat gagacagaag ccagatcacc tttttgggcc tggcttcaga
                                                                         540
 agtcacccag catcacttct gctgcattta tttcttaaaa cacaaatatc aaaccccatc
                                                                         600
 tettgatggg agggggeete atggtttgta aacatgttet aaacteeact etgeeeggee
                                                                         660
 ttggctcaac gtgtctggta atgtgtgggc tgtgaggctc cccgaacgta gacctcagac
                                                                         720
 tgcaacgctg gccgttacag ggtctggcac acgggcccac gtcaggccca tcgccacagt
                                                                         780
 gatggttgtt ctgtgactgt ttctggtggc ctctgctcca cactccaggc tgacgctgtg
                                                                         840
 ccccttccac tgggaccctc gggtggcttc catgcacttg tgccctaaat cctgctccta
                                                                         900
 gactaaactt catctcctgt gttctcattc tgcagcatgg ctgttaggga acctgaccat
                                                                         960
 ctgcagcgcg tctcgttgcc aaggtataat gtcagtgcct cccttcagtg gctcccatgt
                                                                        1020
 cacagaattg teetgeagee etggeacatg tgtgeeatgt gggagetggg geaggteete
                                                                        1080
 tttcatcctg tggctccgag ggagggggcc gctccttccc cagtctctac cctgacttgg
                                                                        1140
 ccctcgtcct gcagccactc agagagcacg atggagctgg agettcagtt ttgaccaaat
                                                                        1200
 gcgtgtcgcc ggcttttgtg tgtgtgtgtg tgtgacagag ccagaccctg tctttaaaaa
                                                                        1260
aaaaaaaaa aaactcgagg gggggccc
                                                                        1288
<210> 32
<211> 3280
 <212> DNA
<213> Homo sapiens
<400> 32
gcgtccggtg agaggcaagg acttttcatc tttagggcat tctgaccacg tcctgcttca
                                                                         60
gagagattgt teceggegte teagtgetat ggggageagg tteeteetgg teetgetete
                                                                        120
aggteteact gtettactgg etetgecagg ateagaagee aagaattetg gagetteetg
                                                                        180
tectecatge cetaaatatg ceagetgeea caacageace cactgtaett gtgaagatgg
                                                                        240
ctttcgggcc aggtctggca ggacatactt tcatgattcc tctgagaagt gtgaagatat
                                                                        300
taatgaatgt gaaaccgggc tggcaaagtg caagtataaa gcatattgta ggaataaagt
                                                                        360
tggaggttac atctgtagct gtttggtaaa atatacttta ttcaactttc tggctggtat
                                                                        420
tatagattat gatcatccgg attgttatga gaacaatagt caagggacga cacagtcaaa
                                                                        480
cgtggatatt tgggtgagtg gggtgaagcc tggatttggg aaacagctgg tacgtataac
                                                                        540
tatgccattt tcctacccaa acattaacat gtcttcctgt gatttttagg gtagggtagt
                                                                        600
totatocagg ggtaattttg toototgtoo caaggtoato tgtcaatgac tggggacact
                                                                        660
tttggttgtc atacettggg ggtgatgtgt gtgactggca tetggtggat ggagaceagg
                                                                        720
gatacagete aacateetae agtgeecagg acageeteee acaateaaga agtgeetagt
                                                                        780
gccatatgtc catagagata gagaaataca agtgtagggg gaaagtgcct cagctggcat
                                                                        840
tcaaagacct acatcaagca cctagattct caatgccaca cgcaccttgt agcaccttaa
                                                                        900
taaatategt tgeettetgg geteeceact ceaacaettg tgeatattee etatttetta
                                                                        960
catttcagta agagtcaatc taattagctt aatttttttg gaaggaaaat ctgagaagaa
                                                                       1020
atggaagcag agaggacttt gcaagaaggg ctactcaact aattcaaagc gtggagttga
                                                                       1080
gcatctggaa tgcgagtttt gcttctccag gaaagggtca aatttctgaa tttgatatag
                                                                       1140
tctatgaaac caagaggtgc aatgagacaa gggagaatgc ttttctggaa gctggaaata
                                                                       1200
acaccatgga tatcaactgt gctgatgctt taaaaggaaa cctaagagag agcactgcag
                                                                       1260
ttgccctatc acttatcaat ctcttgggga tattctgaat gcatcctttt ttagtaaacg
                                                                       1320
aaaagggatg caggaagtaa aactgaactc ttacgttgtg agcggcacca tcggtttgaa
                                                                       1380
ggaaaaaatt teeetetetg aacetgtgtt eetgaetttt egecataate ageetggtga
                                                                       1440
caagagaaca aaacatatct gtgtctactg ggagggatca gagggaggcc gctggtccac
                                                                       1500
ggagggctgc tctcatgtgc acagcaacgg ttcttacacc aaatgcaagt gcttccatct
                                                                       1560
gtccagcttt gccgtcctcg tggctcttgc ccccaaggag gaccctgtgc tgaccgtgat
                                                                       1620
cacccaggtg gggctgacca tctctctgct gtgcctcttc ctggccatcc tcaccttcct
                                                                       1680
cctgtgccgg cccatccaga acaccagcac ctccctccat ctagagetet ccctetgcct
                                                                       1740
ettectggee caectectgt teetgaeggg cateaacaga actgageetg aggtgetgtg
                                                                       1800
ctccatcatt gcagggctgc tgcacttcct ctacctggct tgcttcacct ggatgctcct
                                                                       1860
ggaagggctg cacctcttcc tcaccgtcag gaacctcaag gtggccaact acaccagcac
                                                                       1920
gggcagattc aagaagaggt tcatgtaccc tgtaggctac gggatcccag ctgtgattat
                                                                       1980
tgctgtgtca gcaatagttg gaccccagaa ttatggaaca tttactcact gttggctcaa
                                                                       2040
gcttgataaa ggattcatct ggagcttcat ggggccagta gcagtcatta tcttgataaa
                                                                       2100
```

```
cctggtgttc tacttccaag ttctgtggat tttgagaagc aaactttcct ccctcaataa
                                                                        2160
  agaagtttcc accattcagg acaccagagt catgacattt aaagccattt ctcagctatt
                                                                        2220
  tatcctgggc tgttcttggg gccttggttt ttttatggtt gaagaagtag ggaagacgat
                                                                        2280
  tggatcaatc attgcatact cattcaccat catcaacacc cttcagggag tgttgctctt
                                                                        2340
 tgtggtacac tgtctcctta atcgccaggt tcgaatggaa tataaaaagt ggtttagtgg
                                                                        2400
  gatgcggaaa ggggtagaaa ctgaaagcac tgagatgtct cgctctacta cccaaaccaa
                                                                        2460
  aacggaagaa gtggggaagt cetcagaaat ettteataaa ggaggeaetg cateateate
                                                                       2520
  tgcagagtca accaagcaac cgcagccaca ggttcatctc gtctctgctg cttggctaaa
                                                                       2580
 gatgaactga cetggcaagt gecatggcaa tgacceggaa gttaccecte ettteegttt
                                                                       2640
 gtctacagcg cccctgtggt cacacataga ttggacaaat gccactattt ctagctttcc
                                                                       2700
 tgtgaaaagt ctaggctcat tcacctattt tggcttttta tgttcataga aagaacaaga
                                                                       2760
 cattlgggag aattettaga teeagagtee agtagtgtgg caegtgeaat gaagtgtegg
                                                                       2820
 aaggatgcat tttaaagatg gegggeggga gaagtggatt ttettettge agetaetgee
                                                                       2880
 accttgccag aaactcacta actggcatct ggattcagct catagttccc tttctggcct
                                                                       2940
 ctctgctgta ttttatgctc ccaaagatct tacattaaca ctccacattc acataattca
                                                                       3000
 acaattttca tatggatcag tattaaagag ggtgttgcat tttgcaatac aaaaatgcat
                                                                       3060
 tatcaggtgc tggagaggat gtggagaaat aggaacactt ttacactgtt ggtgggactg
                                                                       3120
 taaactagtt caaccattgt ggaagtcagt gtggcgattc ctcagggatc taggactagg
                                                                       3180
 aataccattt gacacagcta teccattaet gggtatatae eccaaggget ataaatcatg
                                                                       3240
 3280
 <210> 33
 <211> 1297
 <212> DNA
 <213> Homo sapiens
 <400> 33
 ccacgcgtcc ggactgcttt acggacattg gatgaagccg aagcatttag aatggtgcct
                                                                        60
 ggcacacagt tggtgcgtga tatggttaag ctttgtgtcc ccacccacat ctcatcttga
                                                                        120
 atgtgacggt tteccegget cectectgee gecatgtgaa gaaggtegtt getteceett
                                                                        180
caccttccac caccatgatt gccatggatg ctccccactc caaagcagcc ctggacagca
                                                                       240
ttaacgaget geeegagaac atcetgetgg agetgtteac geaegtgeee geeegeeage
                                                                       300
tgctgctgaa ctgccgcctg gtctgcagcc tctggcggga cctcatcgac ctcatgaccc
                                                                       360
tetggaaacg caagtgeetg egagaggget teateaceaa ggaetgggae eageeegtgg
                                                                       420
ccgactggaa aatcttctat ttcctacgga gcctgcatag gaacctcctg cgcaacccgt
                                                                       480
gtgctgaaga ggatatgttt gcatggcaaa ttgatttcaa tggtggggac cgctggaagg
                                                                       540
tggagagcct ccctggagcc cacgggacag attttcctga ccccaaagtc aagaagtatt
                                                                       600
ttgtcacatc ctacgaaatg tgcctcaagt cccagctggt ggaccttgta gccgagggct
                                                                       660
actgggagga gctactagac acattccggc cggacatcgt ggttaaggac tggtttgctg
                                                                       720
ccagagccga ctgtggctgc acctaccaac tcaaagtgca gctggcctcg gctgactact
                                                                       780
tegtgttgge eteettegag ecceeacetg tgaccateca acagtggaac aatgccacat
                                                                       840
ggacagaggt etectacace tteteagact acceeegggg tgteegetac atcetettee
                                                                       900
agcatggggg cagggacacc cagtactggg caggctggta tgggccccga gtcaccaaca
                                                                       960
gcagcattgt cgtcagcccc aagatgacca ggaaccaggc ctcctccgag gctcagcctg
                                                                      1020
ggcagaagca tggacaggag gaggctgccc aatcgcccta ccgagctgtt gtccagattt
                                                                      1080
totgacaget gtocatcetg tgtctgggtc agccagaggt tectccagge aggagetgag
                                                                      1140
catggggtgg gcagtgaggt ccctgtacca gcgactcctg ccccggttca accctaccag
                                                                      1200
cttgtggtaa cttactgtca catagctctg acgttttgtt gtaataaatg ttttcaggcc
                                                                      1260
gggcaaaaa aaaaaaaaa aaaaaaaa aaaaaaa
                                                                      1297
<210> 34
<211> 2184
<212> DNA
<213> Homo sapiens
<400> 34
ggcacgaggc gcccgaggat gtgctgctgg ccgctgctcc tgctgtgggg gctgctcccc
                                                                       60
```

- - -

gggacggcg	g cggggggctc	gggccgaaco	tatccgcaco	ggaccctcct	ggactcggag	120
ggcaagtac	ggctgggctg	gagccagcgg	ggcagccaga	a togoottoog	cctccaggtg	180
cgcactgcag	g gctacgtggg	cttcggcttc	tegeceace	g gggccatgg	gtccgccgac	240
atcgtcgtgg	g gcggggtggc	ccacgggcgg	ccctacctcd	äggattattt	tacaaatgca	300
aatagagagt	tgaaaaaaga	tgctcagcaa	gattaccato	tagaatatgo	catggaaaat	360
agcacacaca	a caataattga	atttaccaga	gagctgcata	catgtgacat	aaatgacaag	420
agtataacgg	g atagcactgt	gagagtgato	tgggcctacc	accatgaaga	tgcaggagaa	480
gctggtccca	agtaccatga	ctccaatagg	ggcaccaaga	gtttgcggtt	attgaatcct	540
gagaaaacta	gtgtgctatc	tacagcctta	ccatactttc	, atctggtaaa	tcaggacgtc	600
cccatcccaa	acaaagatac	aacatattgg	tgccaaatgt	ttaagattco	tgtgttccaa	660
gaaaagcato	: atgtaataaa	ggttgagcca	gtgatacaga	gaggccatga	gagtctggtg	720
caccacatco	: tgctctatca	gtgcagcaac	aactttaacg	acagegttet	ggagtccggc	780
cacgagtgct	atcaccccaa	catgcccgat	gcattcctca	cctgtgaaac	tgtgatttt	840
gcctgggcta	ttggtggaga	gggcttttct	tatccacctc	atgttggatt	atcccttggc	900
actccattag	atccgcatta	tgtgctccta	gaagtccatt	atgataatcc	cacttatgag	960
gaaggcttaa	tagataattc	tggactgagg	ttattttaca	caatggatat	aaggaaatat	1020
gatgctgggg	tgattgaggc	tggcctctgg	gtgagcctct	tccataccat	CCCtccaggg	1080
atgcctgagt	tccagtctga	gggtcactgc	actttggagt	gcctggaaga	gctctggaag	1140
ccgaaaagcc	aagtggaatt	catgtgtttg	ctgttcttct	ccatactcac	ctaactaaca	1200
gagcatcagg	ctgcgtcatt	ttcgaaaagg	gaaggaaatg	aaattacttg	cctatgatga	1260
tgattttgac	ttcaatttcc	aggagtttca	gtatctaaag	gaagaacaaa	caatcttacc	1320
aggagataac	ctaattactg	agtgtcgcta	caacacgaaa	gatagagetg	agatgactto	1380
gggaggacta	agcaccagga	gtgaaatgtg	tctctcatac	cttctttatt	acccaagaat	1440
taatcttact	cgatgtgcaa	gtattccaga	cattatggaa	caacttcagt	tcattggggt	1500
taaggagatc	tacagaccag	tcacgacctg	gcctttcatt	atcaaaagtc	ccaagcaata	1560
taaaaacctt	tctttcatgg	atgctatgaa	taagtttaaa	tggactaaaa	aggaaggtet	1620
ctccttcaac	aagctggtcc	tcagcctgcc	agtgaatgtg	agatgttcca	agacagacaa	1680
tgctgagtgg	tcgattccaa	ggaatgacag	cattacctcc	agatatagaa	agaccctata	1740
aagccagaac	cctttggtgt	gtggcacgtc	ttcttccctc	ttcccctgcc	acagagattt	1800
ctcccatcca	acttgcttgt	ttgccttctg	ctaactcagc	tgcacgctga	gcaccaagag	1860
cttgtgatca	aaattctgtt	ggacttgaca	atgttttcta	tgatctgaac	ctgtcatttg	1920
aagtacaggt	taaagactgt	gtccactttg	ggcatgaaga	gtgtggagac	ttttcttccc	1980
cattttccct	ccctcctttt	tcctttccat	gttacatgag	agacatcaat	caggttctct	2040
tctctttctt	agaaatatct	gatgttatat	atacatggtc	aataaaataa	aactggcctg	2100
acttaagata	accattttaa	aaaattgggc	tgtcatgtgg	gaataaaaga	attettett	2160
tcctaaaaaa	aaaaaaaaa	aaaa				2184
<210> 35						
<211> 949						
<212> DNA						
<213> Homo	sapiens					
-400- 25						
<400> 35					•	•
ggcacgagct	cccattgacc	ttattgctca	gagtgcagta	ttagggcaag	gtttcgccac	60
ractecete	catgaatgta	tttctccctc	ctgccctggg	gacatgggga	gtggcccgtt	120
LUTTTCCCCA	tctagtccca	gaaagatggt	gtttggtttt	ctgttgttgg	atttttttt	180
tanner	tttttgcacc	aaagtggcaa	ctaggtcagt	gttgggggat	caagctggcc	240
reggggtggg	gggccccac	ctgcctctcc	ctggttccca	cagtgttagc	gtccctgaaa	300
ayacaatatt	ctctctaaag	caataagggg	tgacgggccg	gggggagtgt	ttgctgctgc	360

agacaatatt ctctctaaag caataagggg tgacgggccg gggggagtgt ttgctgctgc tggcccccag ctccccttcc ctgccaggtg tgggggagac tcctgttgtg actgaatgta 360 420 accecccac ceetgeegea gecaatgeag gggaaggggg acaetettee tgtetettet 480 ccccagctaa agagactttg gacttagggg gcccatgagc ctggagaggc cttaaccctg 540 tgaggaagta tagggggage ceteteceae ecceatecee ttetgagagt ggtcaatgtt 600 tacaagcccc tgagcccccc tgcccaggga ctcagacccc gttgctgtcc ttccccggcc 660 ccggtcttcc tgggccctcg ctgctcctct gccctttctg gggttggggt gggtgcaggg 720 gtcaccgtgt tccctgtctg ccttgtaccc acagtctccc cgcccctct ccaccctgtg 780 tgacttccct ctcttttacc tgctcctgta aatactccct tctcccaata aaacttggtg 840 900

```
<210> 36
  <211> 3338
  <212> DNA
  <213> Homo sapiens
  <220>
  <221> SITE
  <222> (2861)
  <223> n equals a,t,g, or c
  <220>
 <221> SITE
 <222> (3328)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (3330)
 <223> n equals a,t,g, or c
 <400> 36
 taccgggaca ttctggagtc cgagaagtca acggcgcggt tgctgcggcc gccgcgctcc
 ccggcccgag gcgatggaga acggagcggt gtacagcccc actacggagg aggacccggg
                                                                        60
                                                                       120
 ccccgccaga ggcccccgga gcggcctcgc tgcctacttt ttcatgggcc ggctcccatt
 geteeggege gtteteaagg gettgeaget gttgetgtet etgetggeet teatetgtga
                                                                       180
 agaagttgta tcacaatgta ctttatgtgg aggactttat ttttttgagt ttgtaagctg
                                                                       240
 cagtgccttt cttctgagtc tccttatact gattgtgtat tgcactccat tttatgagag
 agttgatacc acaaaagtaa aatcatcgga tttttatatt actttgggaa caggatgtgt
                                                                       360
                                                                       420
gtttttgttg gcatccatca tttttgtttc cacacatgac aggacttcag ctgagattgc
                                                                       480
 tgcaattgtg tttggattta tagcaagttt tatgttccta cttgacttta tcactatgct
                                                                       540
gtatgaaaaa cgacaggagt cccagctgag aaaacctgaa aataccacta gggctgaagc
cctcactgag ccacttaatg cctaaagact ctggggagca gatgttacct aaggtagtga
                                                                       600
                                                                       660
ccctgcattg tggtgcctga gccctggcag aagctcttgt aaaatttgtt aattgtttaa
                                                                       720
accacttett ttggagagea aggggaaggt caagaaggea gttttateaa tattgtgtea
                                                                       780
gtcaccacaa agtaggccag ataagttaaa aaaaattttt ttttaaataa taattgaaac
                                                                       840
ttatctcaaa tggagatttt ggtgggagga ggagaaaaca attgttttta aatcacacag
ctcaacggtt gataaatgat tctgtcattc tgttacaggt cattctttta ctaggcttag
                                                                       900
                                                                       960
cttccaaatt atgctttata gctgtataaa catcgtgatt atattcatct acttagaaat
                                                                      1020
tgttttattt ttaaattaat ttgcttagct gtttgttttg atgcttagat tatgttctgt
taatgggaat ttaacatatt taagaaacca atatttaaaa tgttggtcta ggttttttc
                                                                      1080
                                                                     1140
cttaacatat attaccagge tttactgtat ttcactcage cttaaatgtt ataatatttt
tggataacgg ttattaattc tttgagacct tcgtatagcc tataaaatgt atgggagatg
                                                                     1200
                                                                     1260
ttggtatttt atgtgtataa aagcaacaat atcagcaact tcgtgtttat actgcacctt
                                                                     1320
ggttgttgat gtcaagtaaa aaaaagattg ttttgtaaca cataaaaaaa tggaagaaac
tgataccaca cctaaggacc aaagataaga aagacttttt gcccaagaca gtgaaagtaa
                                                                     1380
                                                                     1440
ttataaaaac aagctttgac cacttaccaa gtatctgaag agatgagttc atactatgat
                                                                     1500
ttagaaagtg gttcaattcc cctgttggca tatgattatt tttactaaaa ttaatacagc
tetgtgggte tteettagtg ttttetttga agccaatetg tttttttag gacaccagee
                                                                     1560
                                                                     1620
tttggttttt catctgttcg agatgcctct tctctgtctc cttatcagat agaaatggag
                                                                     1680
tcatgtgctg ctgcttcatc tagcagaggt tggcctctgg ctctgacact ttttgtcagt
tgtctttagg tggtcctgaa tcttgggccc ttttgattgt gaatactgtg tagcaggatc
                                                                     1740
                                                                     1800
1860
aaaaaaagt aaatatccag ggaaccctgc ccagactaat actgttggtg gcataagaga
atcaagccat totcaagaga taacttoata accagaatty totgttggot agcagetgto
                                                                     1920
                                                                     1980
acagatagge agggeacttg ggatatgace tttetgteea ggtgatteae agaetagaee
                                                                     2040
tttcttatcc tcctcctaga gttttgactt gggactctag tgttaagatg atgagcccgt
                                                                     2100
```

```
gcatcaggtc cttctgcact ttggtggaag tctcccaggg taggtttcct atttgaaaca
                                                                       2160
 gtggaatcat gtttccagtg ataaagttta atgacctcat ccttttttt ttttctcatc
                                                                       2220
 tgccatttgt gtgtcttaga tgggttttaa ttgcatgaat gtggctaatg tggttctcag
                                                                       2280
 aaattggtca gtatggcccr acatagcttc tgctctgtct tactgactca atacctttag
                                                                       2340
 gatttgtatc agagtttgga tactagtgtt agtggtggtg tcaccactac ttaattggga
                                                                      2400
 gataatgaaa ccaatcatgg atgctgtttt tattgggcat gtcatctaag agaggagaaa
                                                                      2460
 tagctgggtt ttgggtctaa ttatgaataa ggactgattc agaaaacgag tttatggtag
                                                                      2520
 gtagactaaa gtttcacatc agactgtacc attgtgattt agacctatct aaaattcaga
                                                                      2580
 gcatatcatc tgggctacct cagggtcacc acccatgtat tgggcttagt caggattgac
                                                                      2640
 agatacattc tcagctggcc tgtcatataa aacatactgt cattgagctt aagctccgct
                                                                      2700
 tgttctgagg tttcacctcc atgtgtttca ttggtgcaaa agtggatctc ttagttggtc
                                                                      2760
 acttaattct ttcttttca gaaagatagt atgttcactg gtatatttgg tcactcttag
                                                                      2820
 aaccttcctt cacattgttt tttatgggac ccatgaatgg nttagccttt cttttctatt
                                                                      2880
 gtagaaggaa ataaatagga gtaaaaagac cattgtagta aataagttca aggggaactt
                                                                      2940
 gggaccagaa accactgtta tgtacaaaaa aatggcaaat tcaataaact caaatttaaa
                                                                      3000
 ataattttta aattaacagt tatgataaat tttatatttt atacaaatag attgcttaga
                                                                      3060
 atggttctca agaattataa gagaaatgaa ctcacagtac aaaaatttta taattactat
                                                                      3120
acttgtgttt tgtttggggg ctgggaaatg tatttttaca ttgtagccaa tcatttata
                                                                      3180
 tttgtcaatt taaatettat gggtettttt ttttttatet etettgatgt eagatttat
                                                                      3240
3300
gggcggccgc tcgcgaatct agaactangn cccacgcg
                                                                      3338
<210> 37
<211> 1563
<212> DNA
<213> Homo sapiens
<400> 37
cggcacgagg agaaaaggac tcagaccttc cagattggct ccagaaggca gctgctgtca
                                                                        60
gtgattttcc ttacagacca ttgcagccac ctaccttgcg ccatggtccc gagacagcat
                                                                       120
ggcgtcttcc cccacacgga cctgcaggtg gcctcttagg agctttctcg ctgaaggagg
                                                                       180
tcatccctct tagtcacatc tgagcagcac tcacttaggc ctcggccccg agggtcatgc
                                                                       240
gggggcacct ggccgggttc ccggctctgt ctggtctggc atctgtctgc ctttgggcta
                                                                       300
cettttegge acageteest gggeetgtgg etgetacaag etggaeteea geteeeetgg
                                                                      360
gctgcagtgc tgctcgcagt ggtccagaaa agaggctggg cacagcggct cctggctcag
                                                                       420
cagcatcett ggegeagget ggeceaggag eteegtgeag ggtgetgeea gttgacceag
                                                                       480
ccccggctgc actcaatgtc cgagaaccgg gctggctggg gggcctcttt gacggagctt
                                                                      540
tactccaagt cttgctgaat ttcctgagga aaagcactga tgttctcatg gacacaaggg
                                                                      600
aggcagaatc tettgaagta gaataaactg cacteettte cagtetteet tteteagett
                                                                      660
ctcctggacc gccagctcct ccatgcgcct cagacactcc caacgccgca ctgcgggggt
                                                                      720
tectecegee etggeceate teatececet tggetgttga tacagetece etgtgtgeat
                                                                      780
gtcgccctgt ggcagatgct cagggacttc tcagactcca ggatcactcc gagtacactg
                                                                      840
accacacage etgetgeaca aactgeageg ecagetaagg accaggaate tgacattgte
                                                                      900
ggaggtgaag ggateetttg tgacataget tteetecagg aggaceatee actgggtgte
                                                                      960
gggggagett etgeeceaag tageaggagg gagetgteaa ggegtggagt teacaeteaa
                                                                     1020
acacteceag aggaeggeae tetgeatggg aegeeateea geteetttga etgtggaate
                                                                     1080
aagtacatta teagetggee eetggeteet ggetgtgaee teeetteget tgaactgage
                                                                     1140
cttgtgtgta agggcgtatc aagttgtatg ggcttcgccg ctgggtgacc ccaaatgcaa
                                                                     1200
ctccaggaag catgggttac caatgactaa ggttaacgat cagctgggcg cggtgctgac
                                                                     1260
gcctgtaatc ccagcacttt gggaggtcga ggcgaatgga tcacgaggtc aggagttcaa
                                                                     1320
gaccageetg gecaacatgg tgaaaceece atetetaeta aaaatacaaa aattageegg
                                                                     1380
gcgtggtggc gggcgcctgt agtcccagct actcaagagg ctgaggcagg agaatcgcgt
                                                                     1440
gaactcggga ggcgcagctt gcagtgagct gagatctcgc ccgctgcact ccagcctggg
                                                                     1500
cggcagagcg agactccgtc tcaaaaaaaa aaaaaaaaa aaaaaaaact cgagggggg
                                                                     1560
ccc
                                                                     1563
```

```
<211> 1048
  <212> DNA
  <213> Homo sapiens
  <400> 38
  cccgggtcga cccacgcgtc cgacaaaaaa caaggtatgt gtgtgccttg ggatgctttt
                                                                       60
  ctgggcttac cccatctgtg tcttcataga ttctctgtcc tgccagccat gtctttggtc
  cacaggagec aetteceaet teaactetee cacaacetee ceaetgttta etetetteat
                                                                      120
 gccctgtgct cttgccccaa atcccttcac ccaattggga aaactggatg atagataagt
                                                                      180
 agtttcttat ttacatgcat ttctctggaa attaaagttg ctttccagct agtatttcct
                                                                      240
 cacgtcggaa gcatgtgtaa aataacacat ggaaactatt tccacctggc acataagaaa
                                                                      300
 aaaagggtct gcctatattg taactgactt ctacactttg catgttttag tctgacccac
                                                                      360
 aattttttt tttggatagt tgaagtaaaa aacgtggtta caagatgact aaagtttggc
                                                                      420
 ctacaatttc ctgtttggac aacaggaaga attgtgcagt gttttgcaac aggactaatt
                                                                      480
 ctagattctc actgccctta aagataacta gagggaaagg gtcttctctt tcttttcat
                                                                      540
 ttattaaaga tacttacagg ttcttcagaa gaagtgatgg gtcctttagg taatgaaata
                                                                      600
 ggtgttgatg gtgttatggg tgatgatgta actggtggcg gctgtataaa gtcttcatcc
                                                                      660
 ttaaacatca gaactgagag ttgggttact gatttaaaag aaaatgtgta ttaatctaca
                                                                      720
 gtecactace ttgactacca ggaacttgte acaactacet catateetgg cacaggatte
                                                                      780
 aaaaggcaaa atctggaacc taattacgtg tagccaaata ttaagcttcc tcttagatca
                                                                      840
 900
                                                                     960
 1020
 aaaaaaaaa aaaaaaaagg gcggccgc
                                                                     1048
 <210> 39
 <211> 1430
 <212> DNA
 <213> Homo sapiens
<400> 39
cggcacgagg agcagctgag tcccttccct gtctttcact cttctggcat cggtggtttt
acttettega ttgaaccetg etteetegae eeeeetggga ggeegeette tteaggegee
                                                                      60
tecettetet ccaegagete getetgacag etgaggaact ggeaagatee tgetacceag
                                                                     120
agggtgaatg ggtatctttc ccggaataat cctaattttt ctaagggtga agtttgcaac
                                                                     180
ggcggccgtg attgtaagcg gacaccagaa aagtaccact gtaagtcatg agatgtctgg
                                                                     240
tetgaattgg aaaccetttg tatatggegg cettgeetet ategtggetg agtttgggae
                                                                     300
tttccctgtg gaccttacca aaacacgact tcaggttcaa ggccaaagca ttgatgcccg
                                                                     360
tttcaaagag ataaaatata gagggatgtt ccatgcgctg tttcgcatct gtaaagagga
                                                                     420
aggtgtattg getetetatt caggaattge teetgegttg ctaagacaag cateatatgg
                                                                     480
caccattaaa attgggattt accaaagctt gaagcgctta ttcgtagaac gtttagaaga
                                                                     540
                                                                     600
tgaaactctt ttaattaata tgatctgtgg ggtagtgtca ggagtgatat cttccactat
agccaatccc accgatgttc taaagattcg aatgcaggct caaggaagct tgttccaagg
                                                                     660
gagcatgatt ggaagcttta tcgatatata ccaacaagaa ggcaccaggg gtctgtggag
                                                                     720
ggtaagtact cttttcctgc tattatccta cactctcagt tcctacaatt tgcagagaat
                                                                     780
ttttttttat ataaagacat aaaatcgtga attataatcc aaaaactaag gtaagaaact
                                                                     840
cctcatctcc cttgaaaggc ccaaaactta tcattggcct tttatttctg cataatgttt
                                                                     900
ggggattata taggtgggga aagttattac attatttgag atggctgttt cgatcatatt
                                                                     960
cacagtgaat gtagttgttc agtgtatttt tttgcaagtt ctgtactaac acgatgatgt
                                                                    1020
                                                                   1080
atgtetttgt agtgettatg ttcaaaaget gttacegget gtgtgetgtg getcatgeet
gtagtcccag cactttggga ggccaacgcg ggtggatcac ttgaggtcag aagttcaaga
                                                                   1140
                                                                   1200
ccagcctggc caacctggtg aaaccccatc tcaactagaa atacaaaaat tagccaggca
tggtggtgca tggctgtggt cacagctact caggaggctg aggcgggaga attgtttgag
                                                                   1260
ccctggaggt ggaggttgca gtgagccaag atcatgccac tgcactccag cctgggtgac
                                                                   1320
ggggcgagac tctgtctcaa aaaaaaaaaa aaaaaaaaa aaaactcgag
                                                                   1380
                                                                   1430
```

<210> 40 <211> 2103

```
<212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (2101)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (2102)
 <223> n equals a,t,g, or c
 <400> 40
 ttcctcgtag cgagcctagt ggcgggtgtt tgcattgaaa cgtgagcgcg acccgacctt
                                                                        60
 aaagagtggg gagcaaaggg aggacagagc cctttaaaac gaggcgggtg gtgcctgccc
                                                                       120
 ctttaagggc ggggcgtccg gacgactgta tctgagcccc agactgcccc gagtttctgt
                                                                       180
 cgcaggctgc gaggaaaggc ccctaggctg ggtctgggtg cttggcggcg gcggcttcct
                                                                       240
 ccccgctcgt cctccccggg cccagaggca cctcggcttc agtcatgctg agcagagtat
                                                                       300
 ggaagcacct gactacgaat gctatccgtg cgagaacagc tattccacga gaggatccgc
                                                                       360
 gagtgtatta tatcaacact tetgtttgca acactgtaca teetetgeca catetteetg
                                                                       420
 accegettea agaageetge tgagtteace acagggtgte etgggeeggg tetmtgagae
                                                                       480
agtggtgatg ttgatgctcc tcactctgct ggtgctaggt atggtgtggg tggcatcagc
                                                                       540
cattgtggac aagaacaagg ccaacagaga gtcactctat gacttttggg agtactatct
                                                                       600
cccctacctc tactcatgca tctccttcct tggggttctg ctgctcctgg ctgctggaag
                                                                       660
acctggagga gcagctgtac tgctcagcct ttgaggaggc agccctgacc cgcaggatct
                                                                       720
gtaatcctac ttcctgctgg ctgcctttag acatggagct gctacacaga caggtcctgg
                                                                       780
ctctgcagac acagagggtc ctgctgggta tgtggcttcg tagggcttgg gatacctggg
                                                                      840
tttccccaag gagagtagcc cctggttcca ggtgcttgct gacagcctcc catccctgca
                                                                      900
cagagaagag gcggaaggct tcagcctgkc aacggaacct gggctacccc ctggctatgc
                                                                      960
tgtgcttgct ggtgctgacg ggcctgtctg tgctcattgt ggccatccac atcctggagc
                                                                     1020
tgctcatcga tgaggctgcc atgccccgag gcatgcaggg tacctcctta ggccaggtct
                                                                     1080
ccttctccaa gctgggctcc tttggtgccg tcattcaggt tgtactcatc ttttacctaa
                                                                     1140
tggtgtcctc agttgtgggc ttctatagct ctccactctt ccggagcctg cggcccagat
                                                                     1200
ggcacgacac tgccatgacg cagataattg ggaactgtgt ctgtctcctg gtcctaagct
                                                                     1260
cagcacttcc tgtcttctct cgaaccctgg ggctcactcg ctttgacctg ctgggtgact
                                                                     1320
ttggacgctt caactggctg ggcaatttct acattgtgtt cctctacaac gcagcctttg
                                                                     1380
caggeeteae cacaetetgt etggtgaaga cetteaetge agetgtgegg geagagetga
                                                                     1440
teegggeett tgggetggae agaetgeege tgeeegtete eggttteece eaggeateta
                                                                     1500
ggaagaccca gcaccagtga cctccagctg ggggtgggaa ggaaaaaact ggacactgcc
                                                                     1560
atctgctgcc taggcctgga gggaagccca aggctacttg gacctcagga cctggaatct
                                                                     1620
gagagggtgg gtggcagagg ggagcagagc catctgcact attgcataat ctgagccaga
                                                                     1680
gtttgggacc aggacctcct gcttttccat acttaactgt ggcctcagca tggggtaggg
                                                                     1740
ctgggtgact gggtctagcc cctgatccca aatctgttta cacatcaatc tgcctcactg
                                                                     1800
ctgttctggg ccatccccat agccatgttt acatgatttg atgtgcaata gggtggggta
                                                                     1860
ggggcaggga aaggactggg ccagggcagg ctcgggagat agattgtctc ccttgcctct
                                                                     1920
ggcccagcag agcctaagca ctgtgctatc ctggaggggc tttggaccac ctgaaagacc
                                                                     1980
aaggggatag ggaggaggag getteageea teageaataa agttgateee agggtttget
                                                                     2040
2100
nna
                                                                     2103
<210> 41
<211> 2349
<212> DNA
<213> Homo sapiens
togaccoacg ogtoogtota titotgigto igtalgiatg ggalgigtot galligigt
```

```
ctgtggtgta tttctgtgtc tgtatgtatg gggtgtgtgt gattgtgtgt gtgtctggtg
  tatetetgta tgtatggggt gtgtetgtgt gtgattgtgt gtgtgtetgg tgtatetetg
                                                                        120
  tgtctgtatg tatggggtgt gtctgtgtgt gattgtgtgt ctgtgtttat gtgtgtgtgt
                                                                        180
  240
  ctggcaaaac aaaaagcttt cagagagatg tgtggccgtc atgatgtctc agctgcaggc
                                                                        300
  atttttcaaa gttatgtttg atgcagctgg cacctttctt ctcggggaca aatgtgatct
                                                                        360
  ttgatcattg agttgagaat gagtttactc aataatttta aaatgcctcg gtgtgagagg
                                                                        420
  atgggetget agtaaacatt tgtggcaaat atgegttgee tttteeteet tcaagtgeat
                                                                        480
  ggggcagget gettetetge acgtecatet gtgggtcagt etgteetgee acetecettg
                                                                        540
  ccccagggct tagtgccctg tgtggatcac ggctgcccca gggatgctgg ggaaggattc
                                                                       600
  ageceaacaa atagattegt tgaegteete teaacgaatg tgtgetttta aaaatgtett
                                                                       660
  taattttgca ctccatttca tacaacaggt gcacatttaa aacatttatt tttaaacagc
                                                                       720
  tetgtettet cagcactgca gactatttet egeteacttg acaggeagae gttgggacta
                                                                       780
 tgttcttttc ctgccaagaa tccatcagcc ccccaatgcc tgcacctcga gtccagactc
                                                                       840
 cttagcgttg cgttgggacg taattccccg ttagcatccc cagccccttc tccccctgtg
                                                                       900
 agecagectg ggttettte ttttteettt etteeetgtt tteettettt etttttggg
                                                                       960
 tttgtaataa gtaatataca gtcttgtggt taagaaatca aaataataaa gaagatacgg
                                                                      1020
 cttgaggatt cttgctctgc actggtcctt ccgccaggac ctggtgcgtc ccccaagagg
                                                                      1080
 gccttgccat gtgcgagctg agacacatgt tcacactttc ctcctttttt gtacaaaagg
                                                                      1140
 tagcacatgg taaatagegt etgeaegett eteceetege tetgtetteg ggaetaetee
                                                                      1200
 acatecatga agaagcagca teettattet ettgeacagt gggcagtgge tacaetgtat
                                                                      1260
 tractogtot otgotgatgg gractattgt agtggtgraa ttagtaacot agtatacatg
                                                                      1320
 cccttcagcg tgtgtactgg tgtatggatc tcacattccc agaggctgag tgaaagaata
                                                                      1380
 aaggettttt gacagtatte etcaggtact cacacacteg etetgtgage cacgtgatea
                                                                      1440
 tgggagegge tecetgeaga ggtgtetgat gggagagget gatgteettg tetgatggtt
                                                                      1500
 getecaaeet gagaatggtg agetgetete ccaatgeaga ettteateae ttgeatetea
                                                                      1560
 tgcgtgagat tgaactgctt ctcatttatg tttttatctt tgtttgccca tattttagct
                                                                     1620
 gggttgctaa tcttttaatg ttttctaggc gcttttcata aattagggag tttagctgta
                                                                     1680
 gtgatgtgag ttgaaagtat tttttcaaat cgtcattgtc ttgactttcc ttttttttt
                                                                     1740
 tettttgeca ettaaaatat gtagttgaat eteagtattt gggggettet ggattttgag
                                                                     1800
 togtagttga aaaggtotto ctattotgag gttatacagg aattcaccca tttttttcct
                                                                     1860
 aatattttat ttccatttaa acttttgacc cacttggatt gattttgttg tatggtgtca
                                                                     1920
 agtatgggtc caagtttgct tttctgacct ggggatccta acatcgttca tcaaaatctc
                                                                     1980
catgttggcc gggcgtggtg gctcaagcct gtaatcccag cactttggga ggctgaggtc
                                                                     2040
ggtggatcac aaggtcagga gatcgagacc accctggcta acacggtgaa atcctatctc
                                                                     2100
tactaaaaat acaacaaatt ageegggtgt ggtggeagge acetgtagte ceagetaete
                                                                     2160
aggaggetga ggcaggagaa tegettgaac eegggaageg gaggttgeag tgagetgaga
                                                                     2220
2280
                                                                     2340
aaaaagggc
                                                                     2349
<210> 42
<211> 1559
<212> DNA
<213> Homo sapiens
atteatgeea aaacatagge ttteagtgee tattacatat ggettteage tetetetaet
gagggatgta ggagtttatt tetgaggtet gageetettt teetttaett eetttaetet
                                                                      60
ttcctaagcc ttctttataa aaactatgca tgttctattg ttttcctttt tgattccctt
                                                                     120
tettttatta teeccagtag gagtgacttg taatteteat atgttagaaa ggcaggtete
                                                                     180
ctggttgaag aaaagatcca cccaagcaag tcagcagttt aataaatttt tgagggggat
                                                                     240
ctcaaatgtg ggaaggattg ttatataaga caaccaaatg atgacatgag acaataaatg
                                                                     300
ctataggaat tatggaggaa taattagcta tttattttct tggttaggga agagatatta
                                                                     360
ttagttgtag aagtaattac taacttctac attttttatt gtggaaatca aaaatatata
                                                                     420
tatgaaaata aaatgttata attgacttca gtgtcccata aaccagcttc caacaattac
                                                                     480
caaattgtga ccaatcettt acacacatge acaggtgtee etecagtate tgtggggcat
                                                                     540
tggttctagg accacttatg gataccaaca tctatggatg ctcaagtccc tgatataaaa
                                                                     600
tggtggacta tatgcatata accegtgtac gtecegtatt aaccaaatee teectagate
                                                                     660
```

acttacaaca cgtaacacaa tctaaatgcc acgttaacaa ccgtcatact gtattaagcg	700
addaddedg coadaalyt acticicia giccagacge cfffffff tgfgfgfgg	780 840
dedecedge eccedageee cetteaacce cetageacat aggaggetga etgagtga-	900
sog to get geget gegetatet atatacatac agacacacat atatacaca.	960
ogendatate electricad adadattatt atcacageta aacadattac cagtaattae	1020
occarected tataceceggi gitcagatti tetagattigg etectaatti tittagagat	1020
catelydate tydetedate catglactgt aatglittgat aacttaagta coottings	1140
geteetta coacticiti attaaattee tegtaattee tegtaataaa tegattetat	1200
cooldyddic teetyldyte tgaattatgt ggtaattgtt cacatgttee agtgteeteb	1260
tacticity yayiiggiag tragatitag aagettgatt aaattgagat titetetetet	
tagaccacca accidagate atcaactigg atcatttatt teattttact tttastator	1320 1380
egotettag aditactict tadaattitg atttaattit ataatcatgt aaaatgette	1440
tadattica adticagate agcaaaacae aataaaatet atteagagaa ggcaaaaaa	1500
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	1559
	1339
<210> 43	
<211> 1766	
<212> DNA	
<213> Homo sapiens	
<400> 43	
cggcacgagg agcactgaag tattcactac atgaagtata ttttgcactg tggacacaaa	60
	120
og og coloci cytagadat tetergatga tacteactat tateceetet etgetagana	180
type cougadygge atalageate galactacat titteagact coattagac	240
adjustately trygatited graatgagig ggettigeat agaattigga agacgaaaga	300
sanguadae ctygotycty caggitygaa cactogcaac aatagatacg gagtitggaa	360
sadgetgeta agetteetta ggaaaattat ttottteaat atttetges tatgestes	420
bounded the contract of the co	480
gottattet gracticagg adactaagac catcactage agttettatt gaggettet	540
by by the control of	600
agagitacat tictetact cagratator tagtagget	660
to be detected agadedaace cigacigata tagiciccat grafificada tagicagana	720
goodgatat tryteetige ccaaatttea tottgaacte taateeccaa toetataans	780
3333000dd ggggggtt CggatCatgg gggcagatcc ctcacggctt ggtgctgta	840
state and the state of the	900
transport caretters transport grant caretters	960
dataget additioning geotteeta gaagegaage agatgeate accagggtta	1020
Togodaded tycayadely ttagecaatt atacetett tetttagaaa ttaaaaaa	1080
Total cadadiyyaa agaalaaagg tattictta tagcaatgca agaagggaah	1140
databagaty yelelyceat tagtgagaga attgagagge titetetaga tagganana	1200
sold tadadyyadd ttattaatat accorcraft argatosttt natasaan	1260
southerest adjudgeda geataageet aacacaatet ggeatgaat aaaattgaa	1320
s-susuage didigitigge titigetett daaacteaaa teaaateete acetettana	1380
stagettate addatitite catterial tatetatear dateareatt another and	1440
Todadaget tyatgattac adadadaagg gaaggactaa gtaatatagg tagagaana	1500
January and additional literation of the same of the s	1560
	1620
according addition attachment the according to the contract of	1680
on the state of th	1740
aaaaaaaaa actcgagggg gggccc	1766
<210> 44	
<211> 2572	
<211> 25/2 <212> DNA	
<213> Nome anniame	
<213> Homo sapiens	

<220>

==

```
<221> SITE
   <222> (2527)
   <223> n equals a,t,g, or c
  aatteggeae gagtetggae ettettagmt tgettgatat caggttgttt ttgtageeat
  tttgcttcac agtcacctgg aatgccggga gcccctgctc atcccgatcc tctccttgta
                                                                           60
  catgggcgca cttgtgcgct gcaccaccct gtgcctgggc tactacaaga acattcacga
                                                                          120
  catcatccct gacagaagtg gcccggagct ggggggagat gcaacaataa gaaagatgct
                                                                          180
  gagettetgg tggcetttgg etetaattet ggccacacag agaatcagte ggcetattgt
                                                                          240
  caacctcttt gtttcccggg accttggtgg cagttctgca gccacagagg cagtggcgat
                                                                          300
  tttgacagcc acataccctg tggtcacatg ccatacggct ggttgacgga aatccgtgct
                                                                          360
  gtgtatcctg ctttcgacaa gaataacccc agcaacaaac tggtgagcac gagcaacaca
                                                                          420
 gtcacggcag cccacatcaa gaagttcacc ttcgtctgca tggctctgtc actcacgctc
                                                                          480
 tgtttcgtga tgttttggac acccaacgtg tctgagaaaa tcttgataga catcatcgga
                                                                          540
 gtggactttg cetttgcaga actetgtgtt gtteetttge ggatettete ettettecca
                                                                          600
 gttccagtca cagtgagggc gcatctcacc gggtggctga tgacactgaa gaaaaccttc
                                                                         660
 gtccttgccc ccagctctgt gctgcggatc atcgtcctca tcgccagcct cgtggtccta
                                                                         720
 ccctacctgg gggtgcacgg tgcgaccctg ggcgtgggct ccctcctggc gggctttgtg
                                                                         780
 ggagaatcca ccatggtcgc catcgctgcg tgctatgtct accggaagca gaaaaagaag
                                                                         840
 atggagaatg agteggeeac ggagggggaa gaetetgeea tgacagaeat geeteegaea
                                                                         900
 gaggaggtga cagacatcgt ggaaatgaga gaggagaatg aataaggcac gggacgccat
                                                                         960
 gggcactgca gggacagtca gtcaggatga cacttcggca tcatctcttc cctctcccat
                                                                        1020
 cgtattttgt tccctttttt ttgttttgtt ttggtaatga aagaggcctt gatttaaagg
                                                                        1080
 tttcgtgtca attctctagc atactgggta tgctcacact gacgggggga cctagtgaat
                                                                        1140
 ggtctttact gttgctatgt aaaaacaaac gaaacaactg acttcatacc cctgcctcac
                                                                        1200
 gaaaacccaa aagacacagc tgcctcacgg ttgacgttgt gtcctcctcc cctggacaat
                                                                        1260
 ctcctcttgg aaccaaagga ctgcagctgt gccatcgcgc ctcggtcacc ctgcacagca
                                                                        1320
 ggccacagac teteetgtee ceetteateg etettaagaa teaacaggtt aaaactegge
                                                                        1380
 ttcctttgat ttgcttccca gtcacatggc cgtacaaaga gatggagccc cggtggcctc
                                                                        1440
 ttaaatttcc cttccgccac ggagttcgaa accatctact ccacacatgc aggaggcggg
                                                                        1500
 tggcacgetg cagcccggag tccccgttca cactgaggaa cggagacctg tgaccacagc
                                                                       1560
aggetgacag atggacagaa tetecegtag aaaggtttgg tttgaaatge eeegggggea
                                                                       1620
gcaaactgac atggttgaat gatagcattt cactctgcgt tetectagat ctgagcaagc
                                                                       1680
tgtcagttct cacccccacc gtgtatatac atgagctaac ttttttaaat tgtcacaaaa
                                                                       1740
gegeatetee agatteeaga ceetgeegea tgaettttee tgaaggettg etttteeete
                                                                       1800
gcctttcctg aaggtegeat tagagegagt cacatggage atcctaactt tgcattttag
                                                                       1860
tttttacagt gaactgaage tttaagtete atecageatt etaatgeeag gttgetgtag
                                                                       1920
ggtaactttt gaagtagata tattacetgg ttetgetate ettagteata aetetgeggt
                                                                       1980
acaggtaatt gagaatgtac tacggtactt ccctcccaca ccatacgata aagcaagaca
                                                                       2040
ttttataacg ataccagagt cactatgtgg tcctccctga aataacgcat tcgaaatcca
                                                                       2100
tgcagtgcag tatatttttc taagttttgg aaagcaggtt ttttccttta aaaaaattat
                                                                       2160
agacacggtt cactaaattg atttagtcag aattcctaga ctgaaagaac ctaaacaaaa
                                                                       2220
aaatatttta aagatataaa tatatgctgt atatgttatg taatttattt taggctataa
                                                                       2280
tacatttcct attttcgcat tttcaataaa atgtctctaa tacaatacgg tgattgcttg
                                                                       2340
tgtgctcaac atacctgcag ttgaaacgta ttgtatcaat gaacattgta ccttattggc
                                                                       2400
agcagtttta taaagtccgt catttgcatt tgaatgtaag gctcagtaaa tgacagaact
                                                                       2460
atttttncat tatgggtaac tgggggaata aatggggtca ctgggagtag gg
                                                                      2520
                                                                      2572
```

<210> 45

<211> 526

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (66)

<213> Homo sapiens

```
<223> n equals a,t,g, or c
  <220>
  <221> SITE
 <222> (106)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (484)
 <223> n equals a,t,g, or c
 <400> 45
 ctctgacagc ttcctctttg gccaagccct gcctctgtac agcctcgagt ggacagccag
                                                                          60
 aggtcnagac tggagcccag agcccaagat ggagccccag ctgggncctg aggctgccgc
                                                                         120
 cctccgccct ggctggctgg ccctgctgct gtgggtctca gccctgagct gttcttctc
                                                                         180
 cttgccagct tcttcccttt cttctctggt gccccaagtc agaaccagct acaattttgg
                                                                         240
 aaggactttc ctcggtcttg ataaatgcaa tgcctgcatc gggacatcta tttgcaagaa
                                                                         300
 gttctttaaa gaagaaataa gatctgacaa ctggctggct tcccaccttg ggactgcctc
                                                                         360
 ccgattccct ttgstttctt atccttgcaa attactccar atgattycca aaatctggsg
                                                                         420
 sccttgtgga ratcttttaa ctggtcagca awtwtcaaac gaaatctcca aacaggaaat
                                                                         480
 cttntgcctc ctgcatccac ccccaaagaa cttgcacatt gacgtt
                                                                         526
 <210> 46
 <211> 1032
 <212> DNA
 <213> Homo sapiens
 <220>
<221> SITE
<222> (974)
<223> n equals a,t,g, or c
<400> 46
gtaaaattgt caaacattta tttttaaaaa taaaatgtgg tgggcggtta tgggtggagt
                                                                         60
tattgggtcc tggctgtcgc cactgagcat tgctgaatgt tgtcatgact tgtggacttc
                                                                        120
tcagagetge gageatgegg gtgetetttg tggagatttg etetgtgeet gtaggaaggt
                                                                        180
gggtgtgtgg tgtgcattgc agcaacattg gtggaacaga tgtgtgtgcc cccatgctgt
                                                                        240
catcagggtt cattgcacag gagcctcata cacactacag aaaatatgct cttgtaaccc
                                                                        300
caagttcatg gggaggcatc cacacagatg gcagcagatc aggaaatgct ctcagccagt
                                                                        360
cctcaggggc agtagggcag ccttcatctg ggtgcgcttg gcagccctga actttatctc
                                                                        420
ttcttttcgt tgtatttctt taatatctta ttcagcattt tagtttcttt agttggatga
                                                                        480
gttgttcagt gtctttgtac tgtacagttc tcctgtgctg ttggaatgga cgtcatgggc
                                                                        540
tatgttactc ttttgaaata caaatttctc tttagaagtt ttggaaaatc tgttagagga
                                                                        600
ggagttctga gaatatgaga agacttcttg ccacatagtt gtgcctgccc gatgtatttt
                                                                        660
tccaaaaaaa ttgaaggtga acctgttgga agctcagttc tgagtttaat tgactacttt
                                                                        720
gcagagacac aacactctct ttstagcatg tcccttggaa gaaaataggc tggtgcttgg
                                                                        780
gtggtgatca gaactggcag ttattctagg cctaaaacag tttggcaata aaataagaaa
                                                                        840
aaggggccac atgtggtggc tcatgcttta atccccacac tttgagaggc tgaggcaggg
                                                                        900
ggattgcgtg aagccagcct gggcaacata aaatttaaaa attagccaga catggtggca
                                                                        960
tgtgcctgtg ggtncagcaa cctgggtggc taagaccgaa ggatcacttg agcccaggaa
                                                                       1020
gtcaaggctg ta
                                                                       1032
<210> 47
<211> 2680
<212> DNA
```

```
<400> 47
  ggcacgagtc ttacagcaag tacttaacaa acttagattt ccataatcag atttatctca
                                                                         60
  aaccgcaaaa aaaagctgat tttattatgt taaactgata ggagattttt atgagagatg
                                                                        120
  aaagtaagcg attttaactt tttaattttt ttaatttttg ctctgtttct cacacttgag
  gettttetga aatteaccaa aagagttttg geagttgtgg ggaatttace agaaccccc
                                                                        180
                                                                        240
  attattaaaa caattggttt cttatattaa aatcatcaga caggttttgt gtacacagac
                                                                        300
  atttatgtta agccactgga tagaactaaa ctgtgtgtag gaacctgttt tctctaagaa
                                                                        360
  aggactcctt tgagttctgt aggatatgtg cttaaaggtt ttctgtgtca agctattta
                                                                        420
  tgaaataaaa ttttacaaac aatgaagctt cattttcaca cgttgtattt gtttaatgca
                                                                        480
 gacttggcct atacgagttt ttgaaattaa gggttttttc cgtaaaacta gcttaggagt
 ttaatgagtc actgtgtctc tcacctctgt aggatgagca cagagatcac atttgagaag
                                                                       540
                                                                       600
 caggtgtggg tgttttctgc tggtcacgtc tcctctgcct ctcctctggc tcagcagtta
                                                                       660
 cttttcctca gtcatggggg ctgcattcag ctgcgatgcc agcattaggg ccaagctctc
 cogettetgt gacgttttat tacacgtact etttateagt agtagetgee attecetgag
                                                                       720
                                                                       780
 tgttccttcc accgtgcact gagaagggct gaggatttcc gaacatcaac ctcctctaac
                                                                       840
 cctctgtatg ggaattagat ttttatcatc ctgactgttg gggctttcta gagctcaggc
                                                                       900
 aacttgcccc tgggcgggat ggagcctgcc tggggggagt gtgaatgagg cttctgttgt
                                                                       960
 ctgtggcttg gctctctgtg gttggaaact tcagaagcag tagattccag gtgcaacatt
 tggtgcccgg acaagttgta aaacctgacc ttggggaggg ccttcactgg cagccagtgt
                                                                      1020
                                                                      1080
 ccacccgtga gcccggtgag cacagacgtt agagccaggg ccggcgggct cagaggttgg
                                                                      1140
 agccggggac cagtgggctc agaggtcgga gccggggacc agtgggctca gaggtcggag
                                                                      1200
 ccatgtccct gccgtgaagc tctgagggtt ggtggtgaca tcggggcgat tgtggcatcg
                                                                      1260
 cgattgtggt gtcttggctg cctttggccg gcgctgccac tgccctggtt tcatgacgta
 ggcaccacat ggtcctgtgt tgtctcactt acggctccac ttctggcttt tttctacttg
                                                                      1320
                                                                      1380
 ctcaccettg agataaactt ctatttcatt ttcatttcgg tttatatgtg ggttttcttc
                                                                      1440
 caggtctgat gttaagccta taatattgca atgtgatgtt ttgaagttaa ggtgtaatag
                                                                      1500
 agccagtgaa ccaagggttc acaccccagt gaaatacaaa tattcagaat tgagccactg
                                                                      1560
 tgttgccata ctgattatgt aatgtgtgat taacaagtat aatgtgtcac tttcaacatc
                                                                      1620
 agtttcatgc caaagttgca ttttattaga ttatttggga gttcactttg ggcccaaagg
                                                                      1680
 ctcgtgtcta cataataata acttatgatt tttctttttg tctttgttt attttttgtt
 ttgtgttttt tgctttctag accatgccag agtaatctca gctttcttta gttactggat
                                                                      1740
                                                                     1800
cacacatatc cttcctgaga agagcagtga ctaaaatgga atatctcttt aagaacagct
                                                                     1860
cctctttaac aaaaaaactt aaaagacaaa tgtgagatgg gcttagagtt agttctctgg
gaacttgaaa gacatttatg ccatattatt tattcacgtg tttgttcctg gtgggcaaga
                                                                     1920
                                                                     1980
tgccatctga ggcttcagat gagaaattgg ggtaaaatgg aaatttttca cttatttgca
                                                                     2040
attatata tottgaatta ctacataaaa cttgattctg tttctctact tattgtaaaa
                                                                     2100
attgaaaatg gacattctgt taagttaaat gtatagtttg aagctcatat atttttatga
                                                                     2160
agttttgaat caccttgtat ctgaaagtct ctgctttaag aatgctttct gggtattaaa
                                                                     2220
atcatatcag gtacctgttt ttcctgtttc gattttcttt tctgtgatag aagcagtcgt
                                                                     2280
                                                                     2340
cagttettgg tattactaag tgttaaaage atcagtcagg cegggtgegg tggetcaege
                                                                     2400
ctgtaatcct ggcactttgg gagaccgagg caggcggatc acaacgtcag gagatcgaga
gcatcctggc taacacggtg aaaccccatc tctattaaaa atacaaaaaa ttagccgggc
                                                                     2460
                                                                     2520
gtggtggcgg gcgcctgtag tcccagctac tctggaggct gaggcaagag aatggcatga
                                                                     2580
acctgggagg cagagettge agtgagetga gategeacea etgeaetgea geetgggega
                                                                     2640
2680
<210> 48
<211> 1730
<212> DNA
<213> Homo sapiens
<400> 48
cccacgcgtc cgggggcctg tggcggagtc caagaggaac taggccgcct tcgggatggt
                                                                      60
gtggagcgct gctcctgccc cctgttgcct cctcggggtc ctggggctgg tccaggtgtt
gggggcccaa gccgtgggcc cctggacggc ttcagcgtgt ttgggggcag ctcaggctca
                                                                      120
                                                                     180
geeetgeagg ceetgeaagg agagetetet gaggttatte teagetteag eteecteaat
```

gactcactg	a atgageteca	a gaccactgt	g gagggccag	g gcgctgatct	ggctgacctg	300
ggggcaacc	a aggaccgtat	catttctgaq	g attaacaggo	c tgcagcagga	ggccacagag	360
catgctaca	g agagtgaaga	a gegetteega	a ggcctagagg	g agggacaagc	acaggccggc	420
cagtgcccc	a gcttagaggg	g gcgattgggd	cgtcttgagg	g gtgtctgtga	acggttggac	480
actgtggct	g ggggactgca	a gggcctgcgd	gagggccttt	ccagacacgt	ggctgggctc	540
rgggcrggg	c teegggaaac	caacaccacc	agccagatgo	aggcagccct	gctggagaag	600
ctggtcggg	g gacaggcggg	r cctgggcagg	, cggctgggtg	, cccttaacag	ctccctgcag	660
ctcctggag	g accgtctgca	ccageteage	ctgaaggacc	tcactgggcc	tgcaggagag	720
gctgggccc	c cagggeetee	: tgggctgcag	ggacccccag	gccctgctgg	acctccagga	780
tcaccaggca	a aggacgggca	agagggcccc	atcgggccac	caggtcctca	aggtgaacag	840
ggagtggag	g gggcaccagc	agcccctgtg	r ccccaagtgg	cattttcagc	tgctctgagt	900
ttgccccggt	ctgaaccagg	cacggttccc	: ttcgacagag	tcctgctcaa	tgatggaggc	960
tattatgat	c cagagacagg	cgtgttcaca	gtcgccactg	gctggacgct	acttgctgag	1020
cgcggtgctg	g actgggcacc	ggcacgagaa	agtggaggcc	gtgctgtggc	cgctccaacc	1080
agggcgtggc	cggcgtagac	tccgggtggc	tacgagcctg	agggcctgga	gaataagccg	1140
grggccgaga	a gccagcccag	cccgggcacc	ctgggcgtct	tcagcctcat	cctgccgctg	1200
caggeeggg	, acacggtctg	cgtcgacctg	gtcatggggc	agctggcgca	ctcggaggag	1260
ccgctcacca	tcttcagcgg	ggccctgctc	tatggggacc	cagagettga	acacgcgtag	1320
actggggtc	: cgcccgacgt	gtctacgtcg	gctgaagaga	cagcgggggc	ggcgggctcc	1380
rggggtctcg	, cctgagacgg	ggcacctagc	cctgggcgag	cgccgcaccc	gggcccgcag	1440
cggcaccgcg	r cccagagcgg	cctctcccca	cgcccggggc	gcgccggctc	agggaggctc	1500
ggggcgccc	: atgcagactt	ttggcctggc	gcgatccccc	aagaacccct	ccagggccgg	1560
cctgcggagg	agccgatcct	cgcaccctcc	gctccctcca	ctggccctcc	aggtcgattc	1620
cetgggetee	aggctccccc	gcgcgggcgc	cgcccaccgc	catactaaac	gatcgaggaa	1680
taaagacact	tggtttttct	aaaaaaaaa	aaaaaaaaa	aaggggccgg		1730
<211> 1275 <212> DNA <213> Homo <400> 49		•				
	2001001011					
gtactgttta	ageteetet	tastttass	ctccttaaac	tttgccagtg	aaatggagaa	60
Catcottott	tttttaatta	tgetteaag	gttttgctta	tttttgataa	tgctcctgca	120
CCacccctt	attocaacta	atagataaa	acgtcacact	gttgtctctc	tctccaagca	180
tacctgagga	gaectttaa	ataggattata	gagttatage	agcttttaag	gcctgattac	240
Cactagtaga	atteteena	craygerare	getgeaaata	agaaagatcc	tttttaaaca	300
ttaggggtga	actcaccaac	gactatata	cctatggcag	catcaacaac	cttggttaag	360
ttatccatga	Cttcaaaaca	ttactaaga	tanagetateta	ggaagagata	ttgaagacat	420
aggetttagt	traaatraca	aacaacttta	cgaggatgtt	gcaaaaaaa	aatattcaca	480
aggtcacage	tataataact	aacaacttta	agetggatgt	ggaggaggat	gatattgagg	540
aatocataoc	aagaggeace	gayyaattga	Claatgacga	gttactggaa	ccagaacaga	600
cacagtgatg	aagaggaaag	gaaactacag	gagaaaaaa	gaaaagcccc	caaggatatt	660
aaaatottoa	Ctccaacata	gaagettttg	cagacctcaa	caaactcctt	aaaacgtttg	720
tatctattta	Caacacatg	tataataat	cattagtaga	gaggaatgtt	catcgtgcat	780
acacatttct	CaaycadatC	catyatgaat	aaaagaaaca	aaccaaacaa	accaccatga	840
toccanaana	yaaaayaytg	acateteaag	aagagcctca	ggcaggtcct	tcaggaggta	900
ttaaagggc+	tocagtage	yctatcacag	gacatgacag	ctccctgggt	gttattgccc	960
accetacate	cccaycagga	Laaggtgtgg	agatggaagg	tgccgatttt	tatgatcctg	1020
aatotaaaaa	gyuctagget	aatgggtaat	ggtgtgtttg	tgtcttagtt	tttaaagaaa	1080
tataarrarr	gradadata	aaaataacat	taatataggt	aaaagtttat	agaataagga	1140
aaraataaa	yaaaaaaatt	accttcctct	gtatttgtgt	ttcaagctgt (gttactacaa	1200
aagaacaaaa	aagttaagag	aaattgtttt	taaagtaaaa	tagttacaat a	aatctaaaaa	1260
uuuaaaad	aaaaa					1275

<210> 50 <211> 1762

```
<212> DNA
  <213> Homo sapiens
  <220>
  <221> SITE
  <222> (447)
 <223> n equals a,t,g, or c
 <400> 50
 ggtggcctag agatgctgct gccgcggttg cagttgtcgc gcacgcctct gcccgccagc
                                                                           60
 ccgctccacc gccgtagcgc ccgagtgtcg gggggcgcac ccgagtcggg ccatgaggcc
                                                                          120
 gggaaccgcg ctacaggccg tgctgctggc cgtgctgctg gtggggctgc gggccgcgac
                                                                          180
 gggtcgcctg ctgagtgggc agccagtctg ccggggaggg acacagaggc cttgttataa
                                                                          240
 agtcatttac ttccatgata cttctcgaag actgaacttt gaggaagcca aagaagcctg
                                                                          300
 caggagggat ggaggccagc tagtcagcat cgagtctgaa gatgaacaga aactgataga
                                                                          360
 aaakttcatt gaaaacctct tgccatctga tggtgacttc tggattgggc tcaggaggcg
                                                                          420
 tgaggagaaa caaagcaata gcacagnctg ccaggacctt tatgcttgga ctgatggcag
                                                                          480
 catatcacaa tttaggaact ggtatgtgga tgagccgtcc tgcggcagcg aggtctgcgt
                                                                          540
 ggtcatgtac catcagccat cggcacccgc tggcatcgga ggcccctaca tgttccagtg
                                                                          600
 gaatgatgac cggtgcaaca tgaagaacaa tttcatttgc aaatattctg atgagaaacc
                                                                          660
 agcagtteet tetagagaag etgaaggtga ggaaacagag etgacaacae etgtacttee
                                                                          720
 agaagaaaca caggaagaag atgccaaaaa aacatttaaa gaaagtagag aagctgcctt
                                                                          780
 gaatetggee tacateetaa teeceageat teecettete eteeteettg tggteaceae
                                                                         840
 agttgtatgt tgggtttgga tetgtagaaa aagaaaacgg gagcagccag accetagcac
                                                                         900
 aaagaagcaa cacaccatct ggccctctcc tcaccaggga aacagcccgg acctagaggt
                                                                         960
 ctacaatgtc ataagaaaac aaagcgaagc tgacttagct gagacccggc cagacctgaa
                                                                        1020
 gaatatttca ttccgagtgt gttcgggaga agccactccc gatgacatgt cttgtgacta
                                                                        1080
tgacaacatg gctgtgaacc catcagaaag tgggtttgtg actctggtga gcgtggagag
                                                                        1140
 tggatttgtg accaatgaca tttatgagtt ctccccagac caaatgggga ggagtaagga
                                                                        1200
 gtctggatgg gtggaaaatg aaatatatgg ttattaggac atataaaaaa ctgaaactga
                                                                        1260
caacaatgga aaagaaatga taagcaaaat cctcttattt tctataagga aaatacacag
                                                                        1320
 aaggtctatg aacaagctta gatcaggtcc tgtggatgag catgtggtcc ccacgacctc
                                                                        1380
ctgttggacc cccacgtttt ggctgtatcc tttatcccag ccagtcatcc agctcgacct
                                                                        1440
 tatgagaagg taccttgccc aggtctggca catagtagag tctcaataaa tgtcacttgg
                                                                        1500
ttggttgtat ctaactttta agggacagag ctttacctgg cagtgataaa gatgggctgt
                                                                        1560
ggagcttgga aaaccacctc tgttttcctt gctctataca gcagcacata ttatcataca
                                                                        1620
gacagaaaat ccagaatctt ttcaaagccc acatatggta gcacaggytg gcctgtgcat
                                                                        1680
cggcaattct catatctgtt tttttcaaag aataaaatca aataaagagc aggaaaaaaa
                                                                        1740
aaaaaaaaa aaaaaactcg ag
                                                                        1762
<210> 51
<211> 2059
<212> DNA
<213> Homo sapiens
<400> 51
cggcacgagg tttggtctaa gagtgagcag aactgtgttg agatgggagc acatttggtt
                                                                         60
gtgttcaaca cagaagcaga gcagtggaag gaggcatgca tccccagctg ataccttcgg
                                                                        120
ttattgctgt agttttcatc ttacttctcg gtgtctgttt tattgcaagt tgtttggtga
                                                                        180
ctcatcacaa cttttcacgc tgtaagagag gcacaggagt gcacaagtta gagcaccatg
                                                                        240
caaagctcaa atgcatcaaa gagaaatcag aactgaaaag ttgctgaagg gagcacctgg
                                                                        300
aactgttgtc ctattgactg gagaaccttc cagttccaac tgctattttc ctccttactg
                                                                        360
acaacaagac gtgggctgag agttgaaagg aactgttcag ggatgggggc ccatctgatg
                                                                        420
accatcagca cggaagctga gcagaacttt attattcagt ttctggatag acggctttcc
                                                                        480
tatttccttg gacttagaga tgagaatgcc aaaggtcagt ggcgttgggt gggaccagac
                                                                        540
gccatttaac ccacgccaga gtattctggc ataagaatga acccgacaac tytcagggag
                                                                        600
aaaactgtgt tgttctggtc ttataaccaa gataaatggg cctggaatga tgttccttgt
                                                                        660
aactttgaag caagtaggat ttgtaaaata cctggaacaa cattgaacta gaaactcaga
                                                                        720
```

aagtggtcct tgtgatggaa agagaaaaga aaaaccaatt agaataaaggc agaatgtacg

```
780
 tgcgtcattg gaacacagaa aacatgctgg ttcatacagc gtttttagtc ataatggtct
                                                                         840
 tttttatttt gtttgattca ttcgagacaa catgtgtgta tgtgtgtgtg tgtgtgtgta
                                                                         900
 gataatgtgg tttttgtatg gtgtttgatg gaaggaataa tctttctttg ctttcttagt
                                                                         960
 agtatttcaa ggtgtttact tttcaattgg tgtgcactga atgcatgtat ggaagaatag
                                                                        1020
 cgtgaataat gcaatctctt tgtcattttt tcccctttct cagactctta gctcttaaaa
                                                                        1080
 ttcaaaagat gggatattct aactggtagt ggtgcatcat ttttaaccca aatattgcaa
                                                                        1140
 gcactttaaa gatttgaaac cacattttta ttgtttgatg tttcattttc agacttttta
                                                                        1200
 atgtcagtca ttacaattac attgcatgag gaaaattttt ccagaacaac agtgtggaat
                                                                        1260
 agttctgaat tatgctgttc tacagatgga aaaaagtcca aatgccttta aaaatttact
                                                                        1320
 tettaeteea eecaacaegt ttttgeaaag caagaagtet ttgtaagaea eettaaacaa
                                                                        1380
 agtectteaa ttetacagea gaggaaataa aateeecag aageeaaagg geteacette
                                                                        1440
 acattgttag ttcatgacag acccaggtgt gcttcattag agataacata cattcccttt
                                                                        1500
 ggtatcacag gaagttactg gggattactc gacctcatta cttagctaac gactggataa
                                                                        1560
 aatttettaa ttgtttgaag taacattgta ttegtgtttg eattattaat ttgaatagaa
                                                                        1620
 aataatcaca ttttcaaccc atttatacaa attgttaatg tttctttaga gctgtataac
                                                                        1680
 tatagtttga actagcaagg aagttattgt tttgacaacc agaaattatg cttttctggt
                                                                        1740
 gcatgaaaca ctaattgcaa agggcagtca catccaactt taataaaata tggtggtctt
                                                                        1800
 tottaaaatt ttcaatttgc taatttttcc tggaaaccta actttcctac tactactttt
                                                                        1860
 aaaagcagta tttattttgt aaattagcta ctgatttttt gggtttgaaa tactggtgtt
                                                                        1920
 tgattgtgtg gtggtggtgg tggtgatgga tgcgtgtagc tgtgttaaaa gctgatactt
                                                                        1980
 ttagctgaaa tggtgctaaa taaatttgac tgtgtttgcc atattcacta aaaaaaaaa
                                                                        2040
 aaaaaacta gagggggg
                                                                        2059
 <210> 52
 <211> 3282
 <212> DNA
 <213> Homo sapiens
<400> 52
cccacgcgtc cgacttaaaa gagaagcttt agctgccaaa gattgggaaa gggaaaggac
                                                                         60
aaaaaagacc cctgggctac acggcgtagg tgcagggttt cctactgctg ttcttttatg
                                                                        120
ctgggagctg tggctgtaac caactaggaa ataacgtatg cagcagctat ggctgtcaga
                                                                        180
gagttgtgct tcccaagaca aaggcaagtc ctgtttcttt ttcttttttg gggagtgtcc
                                                                        240
ttggcaggtt ctgggtttgg acgttattcg gtgactgagg aaacagagaa aggatccttt
                                                                        300
gtggtcaatc tggcaaagga tctgggacta gcagaggggg agctggctgc aaggggaacc
                                                                        360
agggtggttt ccgatgataa caaacaatac ctgctcctgg attcacatac cgggaatttg
                                                                        420
ctcacaaatg agaaactgga ccgagagaag ctgtgtggcc ctaaagagcc ctgtatgctg
                                                                        480
tatttccaaa ttttaatgga tgatcccttt cagatttacc gggctgagct gagagtcagg
                                                                        540
gatataaatg atcacgcgcc agtatttcag gacaaagaaa cagtcttaaa aatatcagaa
                                                                        600
aatacagctg aagggacagc atttagacta gaaagagcac aggatccaga tggaggactt
                                                                        660
aacggtatcc aaaactacac gatcagcccc aactcttttt tccatattaa tattagtggc
                                                                        720
ggtgatgaag gcatgatata tccagagcta gtgttggaca aagcactgga tcgggaggag
                                                                        780
cagggagage teagettaae eeteacageg etggatggtg ggtetecate eaggtetggg
                                                                        840
acctctactg tacgcatcgt tgtcttggac gtcaatgaca atgccccaca gtttgcccag
                                                                        900
gctctgtatg agacccaggc tccagaaaac agccccattg ggttccttat tgttaaggta
                                                                        960
tgggcagaag atgtagactc tggagtcaac gcggaagtat cctattcatt ttttgatgcc
                                                                       1020
tcagaaaata ttcgaacaac ctttcaaatc aatccttttt ctggggaaat ctttctcaga
                                                                       1080
gaattgcttg attatgagtt agtaaattct tacaaaataa atatacaggc aatggacggt
                                                                       1140
ggaggccttt ctgcaagatg tagggtttta gtggaagtat tggacaccaa tgacaatccc
                                                                       1200
cctgaactga tcgtatcatc attttccaac tctgttgctg agaattctcc tgagacgccg.
                                                                       1260
ctggctgttt ttaagattaa tgacagagac tctggagaaa atggaaagat ggtttgctac
                                                                       1320
attcaagaga atctgccatt cctactaaaa ccttctgtgg agaattttta catcctaatt
                                                                       1380
acagaaggcg cgctggacag agagatcaga gccgagtaca acatcactat caccgtcact
                                                                       1440
gacttgggga cacccaggct gaaaaccgag cacaacataa ccgtgctggt ctccgacgtc
                                                                       1500
aataacaacg cccccgcctt cacccaaacc tcctacaccc tgttcgtccg cgagaacaac
                                                                       1560
ageccegece tgeacategg cagegteage gecacagaca gagaeteggg caccaaegee
                                                                       1620
caggicacci actogotgot geogococaa gaccogoacc igococtogo etceciggio
                                                                       1680
```

tccatcaacg cggacaacgg ccacctgttc gccctcaggt cgctggacta cgaggccctg	1740
and a second distriction of the second contract description of the second contract des	1800
3 3 3 3 3 3 3 3 4 5 Carrier and a contract of the contract of	1860
s-saugud ggcccgcgc cccgcaccqa qccqqtqccc cqqqcqqcca aqcqqqqcb	1920
	1980
and a second sec	2040
and a second dispersion of the second contract of the second contrac	2100
The state of the s	2160
	2220
	2280
ggtgctcctg ttcgtggcgg tgcggctgtg caggaggagc agggcggcct cggtgggtcg	2340
The state of the s	2400
cetyteccag agetaccagt atgaggtgtg tetyacggga ggeeceggga ceagtgagtt	2460
caagttettg aaaccagtta ttteggatat teaggeacag ggeeetggga ggaagggtga	2520
agaaaattcc accttccgaa atagctttgg atttaatatt cagtaaagtc tgtttttagt	2580
ttcatatact tttggtgtgt tacatagcca tgtttctatt agtttacttt taaatctcaa	2640
atttaagtta ttatgcaact tcaagcatta ttttcaagta gtatacccct gtggttttac	2700
aatgtttcat cattttttg cattaataac aactgggttt aatttaatga gtatttttt	2760
ctaaatgata gtgttaaggt tttaattctt tccaactgcc caaggaatta attactatta	2820
tatctcatta cagaaatctg aggttttgat tcatttcaga gcttgcatct catgattcta	2880
atcacttctg tctatagtgt acttgctcta tttaagaagg catatctaca tttccaaact cattctaaca tcctataat tcgtgtttga aaaccatgtc atttattct acatcatgta	2940
tttaaaaaga aatatttet tactactatg etcatgacaa aatgaaacaa agcatattgt	3000
gagcaatact gaacatcaat aataccetta gtttatatac ttattattt atctttaagc	3060
atgctacttt tacttggcca atattttctt atgttaactt ttgctgatgt ataaaacaga	3120
ctatgcctta taattgaaat aaaattataa tctgcctgaa aatgaataaa aataaaacat	3180
tttgaaattt gtgaaaaaaa aaaaaaaaaa aaaaaaaa	3240
	3282
<211> 1860 <212> DNA <213> Homo sapiens	
< 4 00> 53 .	
aatteggeac gageageagg gagaagagga tgatgageat geeaggeece tggeegagte	
cctgctcctg gccattgctg acctgctctt ctgcccggac ttcacggttc agagccaccg	60
gaggagetggt gtgggattag change cactcoctg gacagetgtg aatacatetg	120
salastesse seggettes eteacteece ecagectaac tacatecage atotaca	180
substance transfer tr	240
James garageacea accededate Ecageteett tottecacaa accasa	300
The state of the s	360
SSS TO THE CACAGO CACO CACCOCC CACCOCCAC CACCOCCACCAC CACCOCCACCAC CACCOCCACCACAC CACCOCCACAC CACCOCCACAC CACCOCCACAC CACCOCCACACACA	420
The state of the s	480
as assume accordaged concentration of the manual concentra	540
single the type of type of the type of	600 660
and a second description accepted the contract and	720
	720 780
addaycagcg acceptate categories	840
datacting decided desagaces authorizes asset-	900
January and Annual Annu	960
and a decectage and added to the contractor acceptant	1020
	1020
and a second deligered agendance against the agent age	1140
and a december of the second contraction of	1200
The state of the s	1260
	1320
Distriction occupying the decident occasion occasion canting the same and the same	1380
ccctccage ccacagaaag aacgcaaget tgtgtaceet teatagtage cacattaaaa	1440

aaaaagggtg aatatgaaat t ggctggagtg c attctcatgc c taatttttgt a tcctgacctc a gccactgcat c	tactaatgag cagtggcacc ctcagcctcc atctttagta agatgatatg	atatttaata atctcagctc tgagtaggtg gagacggggt cccacctcag	ttcccttctc actgcaacct ggattacagg tttgccatgt cctcccaaag	acactcagtc ccgcccttg cgtgtgccac tgtccaggct tgctgggatt	ttgtcgcca ggttcaagca cacacctggc ggtctcgaac acaggcatga	1500 1560 1620 1680 1740 1800 1860
<210> 54 <211> 770 <212> DNA <213> Homo s	sapiens					
<400> 54				•		
aattcggcac g	gagccgaggg	gcgaatccat	acatccgtcg	tgtggactcg	ccctgacatc	60
ccctcaacgc t					-	120
cgtttcctcg c						180
gcgccgggga t				-		240
cagagcaacg c gaaacccggc a						300 360
agcgaggccc a						420
cgcctctcca a						480
ctcgccctct g						540
cggcccgagc g					_	600
ggcctgctga a	aagcggtcg	ggagaacaac	aagccgctga	cggacgaacg	tcgcgcgctg	660
ctcgagagta c					tggggtacca	720
ggtaagtgta c	ccaattcgc	cctatagtga	gtcgtattac	aattcactgg		770
<210> 55 <211> 1093 <212> DNA <213> Homo s	apiens					
<400> 55						
cagattcggc a	cgaggtttt	ttttaagatt	tatttttaa	aaattatttt	tgtggacttg	60
ggtatcaatg a					_	120
gtcatgtgtt g	caccgttct	ctgtatgttt	acgtcctttg	gactggcttc	tcccaggatt	180
cttttctgtt t						240
gtaaaaggga t						300
gctgtttact t						360
agtgtataga a						420
aggaactaag t						480
actettagga c						540 600
agttaagtgt t		_			_	660
ccctttggca g		-		_		720
cttccatgcc g						780
tggtctttt t				-		840
tttctgttct c						900
gatgctaaaa g	ctatccagg a	attttgtttc	tgtttgtttt	aaattttgtg	gttccttccc	960
tttcctcccc c						1020
ccactcgcca g		gtcagaaaaa	taaaatatat	tgtttcttam	aaaaaaaaa	1080
aaaaaaaact c	~-					
	ga					1093

<210> 56 <211> 632 <212> DNA

```
<213> Homo sapiens
  <220>
  <221> SITE
  <222> (29)
  <223> n equals a,t,g, or c
  <220>
  <221> SITE
  <222> (46)
  <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (94)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (162)
 <223> n equals a,t,g, or c
 <400> 56
 cgcaattaat gtgagttagc tcactcatna ggcaccccag gctttncact ttatgctttc
                                                                          60
 cggctcgtat gttgtgtgga attgtgacgg atancaattt cacacaggaa acagctatgc
                                                                          120
 catgattacg ccaagctcga aattaaccct cactaaaggg ancaaaagct ggagctccac
                                                                         180
 cgcggtggcg gccgctctag aactagtgga tcccccgggc tgcaggaatt cggcacgaga
                                                                         240
 ctatgtatat atgtttaata tctgtctttt gaaatgcaga aatagtttaa atgtttcttt
                                                                         300
 gtctattttt ctttttttt aatgctaccc agggaaatat tttcatatca tttttaagtg
                                                                         360
 gcctgcctca atgtatattt atttcttttg aaacaaaaag gttctggaaa ctgttttct
                                                                         420
 gtagctttaa atgaataggt gagcaaaatc tatatgggat gtaatttttt tgttcagtct
                                                                         480
 cttaaaaaat actttgtttt ggtacatttg gttgtgcttg tggggaaaat aaaaacgcag
                                                                         540
agateettat atatttatgt taaagtaata ttttattate tacataaaac agaaatgcac
                                                                         600
aataaaaaa aaaaaaaaa gctcgagggg gg
                                                                         632
<210> 57
<211> 2687
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (1614)
<223> n equals a,t,g, or c
<400> 57
gtacaccatg ggcctccacc tecgececta eegtgtgggg etgeteeegg atggceteet
                                                                         60
gtteetettg etgetgetaa tgetgetege ggaeecageg eteeeggeeg gaegteaece
                                                                         120
cccagtggtg ctggtccctg gtgatttggg taaccaactg gaagccaagc tggacaagcc
                                                                         180
gacagtggtg cactacetet getecaagaa gacegaaage taetteacaa tetggetgaa
                                                                         240
cotggaactg ctgctgcctg tgcatcattg actgctggat tgacaatatc aggctggttt
acaacaaaac atccagggcc acccagtttc ctgatggtgt ggatgtacgt gtccctggct
                                                                        300
                                                                        360
ttgggaagac cttctcactg gagttcctgg accccagcaa aagcagcgtg ggttcctatt
                                                                        420
tccacaccat ggtggagagc cttgtgggct ggggctacac acggggtgag gatgtccgag
                                                                        480
gggctcccta tgactggcgc cgagccccaa atgaaaacgg gccctacttc ctggccctcc
                                                                        540
gcgagatgat cgaggagatg taccagctgt atgggggccc cgtggtgctg gttgcccaca
                                                                        600
gtatgggcaa catgtacacg ctctactttc tgcagcggca gccgcaggcc tggaargaca
                                                                        660
```

```
agtatatccg ggccttcgtg tcactgggtg cgccctgggg gggcgtggcc aagaccctgc
   gcgtcctggc ttcaggagac aacaaccgga tcccagtcat cgggcccctg aagatccggg
                                                                           720
   agcagcagcg gtcagctgtc tccaccagct ggctgctgcc ctacaactac acatggtcac
                                                                           780
   ctgagaaggt gttcgtgcag acacccacaa tcaactacac actgcgggac taccgcaagt
                                                                           840
   tettecagga categgettt gaagatgget ggeteatgeg geaggacaca gaagggetgg
                                                                           900
   tggaagccac gatgccacct ggcgtgcagc tgcactgcct ctatggtact ggcgtcccca
                                                                          960
  caccagacte ettetaetat gagagettee etgacegtga ecetaaaate tgetttggtg
                                                                         1020
  acggcgatgg tactgtgaac ttgaagagtg ccctgcagtg ccaggcctgg cagagccgcc
                                                                         1080
  aggagcacca agtgttgctg caggagctgc caggcagcga gcacatcgag atgctggcca
                                                                         1140
  acgccaccac cetggeetat etgaaacgtg tgeteettgg geeetgaete etgtgeeaca
                                                                         1200
  ggacteetgt ggeteggeeg tggacetget gttggeetet ggggetgtea tggeecaege
                                                                         1260
  gttttgcaaa gtttgtgact caccattcaa ggccccgagt cttggactgt gaagcatctg
                                                                         1320
  ccatggggaa gtgctgtttg ttatcctttc tctgtggcag tgaagaagga agaaatgaga
                                                                         1380
  gtctagactc aagggacact ggatggcaag aatgctgctg atggtggaac tgctgtracc
                                                                         1440
  ttaggaetgg etceaeaggg tggaetgget gggeeetggt eeeagteeet geetggggee
                                                                         1500
  atgtgteece cetatteetg tgggetttte atacttgeet actgggeect ggenesgeag
                                                                         1560
  cettectatg agggatgtta etgggetgtg gteetgtace cagaggteec agggategge
                                                                         1620
  tectggeece tegggtgace etteccacae accagecaca gataggeetg ceaetggtea
                                                                         1680
  tgggtageta gagetgetgg ettecetgtg gettagetgg tggecageet gaetggette
                                                                         1740
  ctgggcgagc ctagtagctc ctgcaggcag gggcagtttg ttgcgttctt cgtggttccc
                                                                         1800
  aggeeetggg acateteact ceaetectae etceettace accaggagea tteaagetet
                                                                         1860
  ggattgggca gcagatgtgc ccccagtccc gcagctgtgt tccaggggcc ctgatttcct
                                                                         1920
  cggatgtgct attggcccca ggactgaagc tgcctccctt caccctggga ctgtggttcc
                                                                        1980
 aaggatgaga gcaggggttg gagccatggc cttctgggaa cctatggaga aagggaatcc
                                                                        2040
 aaggaagcag ccaaggctgc tegeagette eetgagetge acetettget aaceceacea
                                                                        2100
 tracactgre accetgreet agggtetrac tagtarrag tgggtragra ragggetgag
                                                                        2160
 gatggggctc ctatccaccc tggccagcac ccagcttagt gctgggacta gcccagaaac
                                                                        2220
 ttgaatggga ccctgagaga gccaggggtc ccctgaggcc cccctagggg ctttctgtct
                                                                        2280
 gecceagggt getecatgga tetecetgtg geageaggea tggagagtea gggetgeett
                                                                        2340
 catggcagta ggctctaagt gggtgactgg ccacaggccg agaaaagggt acagcctcta
                                                                        2400
 ggtggggttc ccaaagacgc cttcasgctg gactgagctg ctctcccaca gggtttctgt
                                                                        2460
 gcagctggat tttctctgtt gcatacatgc ctggcatctg tctccccttg ttcctgagtg
                                                                        2520
 gccccacatg gggctctgag caggctgtat ctggattctg gcaataaaag tactctggat
                                                                        2580
 gctgtaaaaa aaaaaaaaa aaaaaaaaa ggcggcc
                                                                        2640
                                                                        2687
 <210> 58
 <211> 619
 <212> DNA
 <213> Homo sapiens
 <220>
<221> SITE
 <222> (526)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (619)
<223> n equals a,t,g, or c
<400> 58
teateagett teacetggat tategtgaca geeteetact gettetetat catgtggeca
gagctatett cetaaaatge attgeatagt tgateaagte actetetgge etaaaacett
                                                                         60
cettggetee etgetgeeet caggataaag tetggaeeee teageatgge ttgtgagaet
                                                                        120
catggtgtcc ttgtccctgc tcacctctct ggtctcatca cttgccttct tgcattctgg
                                                                        180
gtcccagcct cctgtatcca gagatgcagt ggctctccat tgccactctg attcctcctt
                                                                        240
tettttggte acagagaaag ggtactttet etgteaaate teaaettaga ettgaettee
                                                                       300
tocaaggage tttggetata eteteteete eegaceeeca eeetggeata etacacagat
                                                                       360
```

```
cactctgggc tcacttgcct gcctaatggt catctcccca gtagactgta agctccttga
  gggcaaggat tgtgttggaa tttttgtatt aacagtgcct ggcttngtgc ctggcaccta
                                                                         480
  540
                                                                         600
  aaaaaaaaa aaaaaaaan
                                                                        619
  <210> 59
  <211> 1378
  <212> DNA
  <213> Homo sapiens
  <400> 59
  togacttotg ggatgtagat totootttat tgaattotto attttcacto gcagtatgat
  ttattttttt cacactgtat tttacactgc atatgtttct cttataacga gggtcacttt
                                                                         60
  cttcctgccc aggattggaa tcaattacag ctatgcaatg ttttttgttt agcatcttct
                                                                        120
  taattacagg cttggcagaa gagttttgtg aacaactctc catttcactg gctgaagaag
                                                                        180
 aaatacaact gtcctctaca gttgaacatt tctgtatgac agcattttca tggtaatgag
                                                                        240
 gattaggttc atactttggt ttttccaaac cggggaaaca aacaacagcc aaaaaccagt
                                                                        300
 gtgcagcttc attaaggggt acaaaaataa aatccttctc aaaaatatct acgtgccggg
                                                                        360
 tocatgtttt tactotocca tgccgttttt gctgtattga cagattagtt gtttcatgat
                                                                        420
 tteteetete tetetgatta aggegtttat agaaaaaaga actgaatata tgaatteggt
                                                                        480
 cagogtotto ottottoagt tittoaagoa ocaagtatit caaataaaag totataataa
                                                                        540
 catcatttaa aaatteteet teatttagae agtgeaggte eteattggta acagagatge
                                                                        600
 ctcccttagc tggaggtggt ggatatacta tcaacttttc tactgggcca atgaagatgg
                                                                        660
 tgtggttttc tccagtttct tcttcttcat caaaaaactg aaattcttgt ttgcttctaa
                                                                       720
 gitgtatttt agattcaaat gatacagttt taattttgtt tteettttge ccacaactte
                                                                       780
 ctttgatgct ctcttcatag gttcttgtac aggcaacaag tctgccatta gcttcttcaa
                                                                       840
 agggaatttt cgcaaaaaa ttggagatgt tattetttat accaatttca ttaatgatac
                                                                       900
 titcaaatac catatttgcc ggaggatcaa ggccattttg aaaaattaaa attatatatt
                                                                       960
 gttcttctaa atttgttaat ttatttactc ctttacaatc attccaaact ttatcctcct
                                                                      1020
 tattcatttg cagttggatg ctcagctttt gataaactgc tggaattgct tgaagaaaca
                                                                      1080
 ctacaggtaa ttttcggaca ttacaccatt cacatttagt tagatcagag gtatttaata
                                                                      1140
 taatototac aggatoatgg totggttogt otagotgtat ottgataaaa totaaacaaa
                                                                      1200
 aaattacagg ctctattaac agccggaaga gtgttcctac tcgtcgacgc ggccgcgaat
                                                                      1260
tcccgggtcg acgagetcae tagtcggcgg ccgetctaga ggatecetcg aggggece
                                                                      1320
                                                                      1378
<210> 60
<211> 1126
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (21)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (35)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (49)
<223> n equals a,t,g, or c
<220>
<221> SITE
```

=====

```
<222> (99)
  <223> n equals a,t,g, or c
  <220>
  <221> SITE
  <222> (1012)
  <223> n equals a,t,g, or c
  <400> 60
  tccggcttcg taatgtggtg ngaaattgta acggnataac aatttcacnc aggaaccagc
  tatgeceatg attacgecaa gttegaaatt aaceeteant aaaggaceaa aagtgaagtt
                                                                        60
  ccaccgcggt gccggcccgc tctagaaact agtggatccc ccgggctgca ggaattcggc
                                                                       120
  acgageegge cegtacegee aggegatege getgatggeg gegetggeag cageggeeaa
                                                                       180
  gaaggtgtgg agegegege ggetgetggt getgetgtte aegeegeteg egetgetgee
                                                                       240
  ggtggtette geeeteeege ceaaggaagg cegetgettg tttgtcatee tgeteatgge
                                                                       300
 ggtgtactgg tgcacggagg ccctgccgct ctcagtgacg gcgctgctgc ccatcgtcct
                                                                       360
 gttccccttc atgggcatct tgccctccaa caaggtctgc ccccagtact tcctcgacac
                                                                       420
 caactteete tteeteagtg ggetgateat ggeeagegee attgaggagt ggaacetgea
                                                                       480
 ccggcgaatc gccctcaaga tcctgatgct tgttggagtc cagccggcca ggctcatcct
                                                                       540
 ggggatgatg gtgaccacct cgttcttgtc catgtggctg agcaacaccg cctccactgc
                                                                       600
 catgatgett cecattgeca atgecateet gaaaagtete titggecaga aggaggtteg
                                                                       660
 aaaggacccc agccaggaga gtgaagagaa cacaggaata gaacccaata ctttcctctc
                                                                       720
 tgaggaaagg ctgaaacttc aagctcccct tgtgataaga cttggtcaga taactgagtc
                                                                       780
 tggtcaatgg aatatgagtg gaaatgatgt gtgcaacttc cgggttctgt ccttcctgcc
                                                                      840
 gggtggaatg tgaatatgat ggcacctggg acccaaagac aggagccaca tcttgagaga
                                                                      900
 tagatggcag atctgcccct gtggctttgg atcatttacc tcagtgaaca cnacaagcat
                                                                      960
 1020
 1080
                                                                     1126
 <210> 61
 <211> 2078
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (337)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (492)
<223> n equals a,t,g, or c
<400> 61
gattattttc aatagattgt caataaatta tcaaattatg gaaatatcct ttagattcct
gtgataatgc tttgtgcatg ttttctatta tttagaaaat gaaaacagca gtcatttgac
                                                                      60
cagccatgtt atgacaccag ataatattct gcaaggatct tccaaacagg gagaaactgc
                                                                     120
aatgggaaca atctttggat atctacattg tgtcaaatgt tatgtgcttt atttcatttt
                                                                     180
catacttata acagetgtat accacagttt ttattateca cattatagag geaaggeact
                                                                     240
gatatcagga acttgagtaa ttggcttaca cagcaantgt agttgaaata gaagccagga
                                                                     300
ctgtctgatt cccaagacag tgttcttgta ctgcaggtaa gcctggtgtt gctgggkttc
                                                                     360
cctagcactg agtattgcag aaatacggag atttatttac tcaggtgaat gatgaagcag
                                                                     420
aatctttatt tngcttttta aattatkgat gggkcckgag ttaagctagt tcaccaatga
                                                                     480
gtctcaaagg gtgaagaact ggcctgtgta gttttaaaac aaagcagcag ttttcttcat
                                                                     540
ttgtagatat tatatttaat ttaccttgca tttccaatag cattaaatat tagaagcgta
                                                                     600
taagcttttt ctttttatgc ctaggggata tggcttaaca ctgagaaaca cttgagaaac
                                                                     660
actgccacat tctaaggaac aaccaataat aagaaaaatt atttcatata aaaggatgca
                                                                     720
                                                                     780
```

```
gcttgcatat aaaatgctga attcaactga agttcattat gctgccttta tgtgagctca
  ggggtgacct tcatgctctg gactcttaaa ggatccagtg tcggatacag ctctgccatt
                                                                          840
  ggagcaatgt gttccataac tagggttgcc taagtatatc ccttcttctg ktatctatag
                                                                          900
                                                                          960
  tttttcttta aaggtgtgct tttcaaaaca tgatgttatt tcatatacmt atatatatat
  atgcacacat acatatacmt catgcatgka tcatatatgt gtgtgtatgt atgtatctat
                                                                         1020
  cttcaatctt tcaggaaagt ttgtaaaawa tagccctaag tagaattatg gggaaacagt
                                                                         1080
  tgcagaactg ctgaaattgg ctcagatgaa aaggaaagag caaaaatcta ggttattcac
                                                                         1140
  aggaagttag tccagctgtg ttttgcgagg catgaaccag gctgaatcag agcaagaaca
                                                                         1200
  ggttgcaatc ttaggtcaca tggcaattaa attccagcta tagaatattg aagacatgta
                                                                         1260
  agtgcttaat ttaaggatta ttcaatgcag tgaataaatt tatatcattc ccaaaagtat
                                                                         1320
  gcacatcaga taatgagtta actgcatgta acttgaaaac cttaaatatt taactgaaag
                                                                         1380
  gattttgtgc atagaaagat gtcttagtca atgtaaaaat gccagtgatt tgtgtagcat
                                                                         1440
  gteettiget aaaagteatt atataaactg aagteactta tettigitte caatgigata
                                                                         1500
                                                                         1560
  atattccaca tttaacaggt gataattatt gaaaatatta gtttgcctca ctaatattgt
  gtccgagata ttctagttta aaccaagtaa ctaatactgt aaaagaagta ttcctttat
                                                                         1620
  gaagccagca ttatcctgat actagaactg ggaagagaca cacacaaaaa gaaaacttca
                                                                         1680
 ggccaatatc gctgatgaac atcaatgtga aaatcttcaa taaaatactg gcaaaccgaa
                                                                         1740
  tecageagea tgaggegggt ggateacaag gteaggagat egagaceate etggetaaca
                                                                         1800
 tggtgaaacc ccatctctac taaaaataca aaaaattagc cgggcatggt gatgggcacc
                                                                         1860
 tgtagtccca gctactgggg aggctgaggc aggagaatgg cgtgaacctg ggaggcggaa
                                                                        1920
 cttacagtga gccgagattg caccactgca ctccagcctg ggtgacagag caagactctg
                                                                        1980
 tctcaaaaaa aaaaaaaaa aaaaaaaaa aactcgta
                                                                        2040
                                                                        2078
 <210> 62
 <211> 762
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (10)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (12)
 <223> n equals a,t,g, or c
<220>
 <221> SITE
 <222> (42)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (219)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (747)
<223> n equals a,t,g, or c
<400> 62
cgagtcagtn ancgagaaag cggaagagcg cccaatcgca anccgcctct ccccgcgcgt
tgcccgattc attaatgcag ctggcacgac aggtttcccg actgaaaagc ggccagtgag
                                                                        60
cgcaacgcaa ttaatgtgag ttagctcact cattaggcac cccaggcttt acacttatg
                                                                       120
cttccggctc gtatgttgtg tggaattgtg agcggatanc aatttcacac aggaaacagc
                                                                       180
```

acgagcacaa tttatatttc gattttttcc agccaccgcc ttgcaaatgg gscctgggat	g giggeggeeg a atgeetatad c teatggitte c ceeettitet c teeageeaca g taageittie	g ctctagaac aaaactttt g gtttcttt gatctgtaa aatactttt atgttgtgta	t agtggatco t tocatgtco t cttgtctto c ttgttttgt t agtcgccat a ttctttcco m tmtgcaatc	c ccgggctgc t caatacttt a tttttctga a tttggatga a agcataatc a tgattagag t gccacatta	a aaagctggag a ggaattcggc g tatcagatta a ggtaaaaggg t tacaggcgtg a agaaatcatt a ggtaaaaagt t ttctttaaaa	300 360 420 480 540 600 660 720 762
<210> 63 <211> 1094 <212> DNA <213> Homo			·			
<400> 63						
ctggacagcg cctgctttcc ccttggtggc ctgccactgc gtctcgctga	ggcccttcc tcttgccggt tgcccctcc cagtggagtg agggcttggg ctttttttat	ccacccaccc cacctttggc ctcctctgtg gacaccagct cggctgacat acagttttgt	tgtggaaagc catctcccgg cctgcctgcc gcgactgcgt gggacgaggc	cogtotggtt tgcgtggttc tctgctgtgc gggagggct ttgcacagct	tgaaagtaga tggtgtcctc agatgtgggt cgggccagtg ggcattgccg gccagctcct	60 120 180 240 300 360 420
cttgccgttc	aggctgctcc	tcatagatga	acaaggccct	accccatato	cttaccetrr	480
ccaggccatg	tttcaagtgc	tccatggcca	catgtggctg	gtggacagtg	cctcagccgc cagctctaga	540 600
acattccatc	accacagagg	gttctgctgg	acagtggcct	tgggggctgt	tttgagggtc	660
ttttacctcc	tctcctggca aagccactat	gtgcgtatgc	totgccattg	tcaagttaca	gttattttcc	720
cctgtcacgg	agggaagggt	gccacaggcc	cagctcctgg	agggggtttg	gatgtctggg	780 840
cggggggaag	ggtgccacag	gcccagctcc	cggaggggt	ttggatgtct	agatagagac	900
tattccatat	cagttccagc	tcccggaggg	ggtttggatg	tctgggcggg	gctttcttca	960
ggagtgttgt agggggggcc	atgataacgg tttcgctaat cggt	ataaatgttt	gaatagctaa	agaggcatac aaaaaaaaaa	aaatagtgtt aaaaaactcg	1020 1080 1094
<210> 64 <211> 1361 <212> DNA <213> Homo	sapiens					
<400> 64						
cccgggtcga	cccacgcgtc	cgaatttctt	tagtgtattt	tatgtacttt	tgtattaatt	60
acattgggta	gactgttcca	taaatgttga	atagatccta	tttattaatt	aatgctattg	120
ttgtgtatcc	Ctataattet	ctgtgtagtt	ggtctatcag	ttgttgggag	aagagtgttg	180
aagtctacaa tcacatattt	tccagctctt	gtatttgtct	attactcctt	tcagttcttt	cagttccatt	240
tagaccattt	tataatttta	tgatgtacct	ctcttgtcac	tttctttact	Ctcaactcta	300
cctgatatta a	atatagtcac ·	tctagctttc	tttggagtaa	tattagtata	atatatettr	360 420
tttcgtcctt	ttacttttaa (ccctatctat	atcattatat	tattttaagt	gagttttffg	480
tagatagcac	ttacttgtgt (ctgtcatgtt	tttaaatcca	ctgtacaagt	acaaatctot	540
ctttattttt d	totoatoto	cttattttga	gacagagtct	cactttgaca	ctcagcctag	600
aatgcagtgg (ctcccaadta (geteactgca deteacaments	accicigeet caddeatdaa	ccgagggtca	agcaatcctt	660
ttttttttt 1	tttttttgta	gagatgqqqa	tttaccator	tactalages	agccctaaec	720 780
ccctgggctc a	aaacgacctg o	ccageteag -	cctcccagag .	aaatatgatt	ttaattagga	840
catttagatc a	atttactttc a	aatggaatta	ttgatgtatt	acaattcaag	tctgtcattt	900

```
tatttattgc cttccatttt tccctccgtt tttcatttct gtgctttttt tctttgcatg
  taaacatgtt ttaggattct gttttgatgt atttagagta ttttttagtg tatcggttta
                                                                        960
                                                                       1020
  tgtagctttt ctagtggatc ctttaagtat tacattatat atacataact aagaagtata
  tcagtgtcat cattttacca gttcaagtaa agtacagaaa tctttgtccc ttttacctgt
                                                                       1080
  cotgettata atgtaaatgt ettaaagatt teettgacae acatttagaa ceatateaca
                                                                       1140
  gtgttgtttt tgcttcagtt gtgaaatata atttagaaac tcatgagaat aaaaacccat
                                                                       1200
  1260
  aaaaaaaaa agggcggccg c
                                                                       1320
                                                                       1361
  <210> 65
  <211> 947
  <212> DNA
  <213> Homo sapiens
  <220>
  <221> SITE
  <222> (67)
  <223> n equals a,t,g, or c
 ctttctatag ttaaagccgg tacgcctgca ggtaccggtc cggaattccc gggtcgaccc
 acgcgtncgg tttctgggtt ctaggccctg cttccttgcc cctttgctgc agaagggcag
                                                                        60
 ctgaaggete accetagaaa cegggeetgg tgggtettae eeggeteact eceteeettg
                                                                       120
 teettacaca tacaggaaga caagacetga gtggtgetgt etttgtgtee gtegtgtatg
                                                                       180
 getetecetg tetreattte treteactet greteraaac etetetet etecettece
                                                                       240
 ectcagtact tartetacag acetatgtge gtgtycetat cettetgtee ttttetetet
                                                                       300
 tragetetee etgeetetea cacacaattt tacatgeeee gaggageeaa gtttgggaca
                                                                       360
 tttaccetce aggeatetgt gteecetett gaagagaaaa cacacagett cacacateca
                                                                       420
 ggcatagggg gcaagctctt ggggcatcag gaccctggag caccaggtcc ttcctggaat
                                                                       480
 attagateca ectggageae caggtetete taagteteae etggggaatt eggteecaee
                                                                       540
 tgggtcacca gttcccacct agagcactgt gtcctgccct agagcacaaa gacctgctcc
                                                                       600
 tecegagaet etetetgaet geageeagge atagtaceet tgeetgtgtt tgeteeetgg
                                                                      660
 tccacagatt tggtggctgg gcaggtgcct ggacagtgat gaggtcttgc cgccttaact
                                                                       720
 gtcccccca gtcacttctc ccacaggccc agcaggacgc agtcctgagg atcagggatt
                                                                      780
 ctacagetge attaaaatea ateetateea aaaaaaaaaa aaaaagggeg geegetetag
                                                                      840
                                                                      900
aggatccaag cttacgtacg cgtgcatgcg acatcaagct ctgaaga
                                                                      947
 <210> 66
<211> 1376
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (18)
<223> n equals a,t,g, or c
<400> 66
ggcgggcggc gcggaagngg cggckgcgcg gccggggcag ccatgtcgcc attgtctgcg
gcgcgggcgg ccctgcgggt ctacgcggta ggcgccgcgg tgatcctggc gcagctgctg
                                                                       60
eggegetgee gegggggett cetggageca gttyteecee cacgacetga eegtgteget
                                                                      120
atagtgacgg gagggacaga tggcattggc tattctacag cgaagcatct ggcgagactt
                                                                      180
ggcatgcatg ttatcatagc tggaaataat gacagcaaag ccaaacaagt tgtaagcaaa
                                                                      240
ataaaagaag aaaccttgaa cgacaaagtg gaatttttat actgtgactt ggcttccatg
                                                                     300
acttecatec ggcagtttgt gcagaagttc aagatgaaga agatteetet ccatgteetg
                                                                     360
atcaacaatg ctggggtgat gatggtccct cagaggaaaa ccagagatgg attcgaagaa
                                                                     420
cattleggee tgaactacet agggeactte etgetgacea acettetett ggatacgetg
                                                                     480
                                                                     540
```

```
aaagagtctg ggtcccctgg ccacagtgcg agggtggtca ccgtctcctc tgccacccat
                                                                        600
  tacgtcgctg agctgaacat ggatgacctt cagagcagtg cctgctactc accccacgca
                                                                        660
  gcctacgccc agagcaagct ggcccttgtc ctgttcacct accacctcca gcggctgctg
                                                                        720
  gcggctgagg gaagccacgt gaccgccaac gtggtggacc ccggggtggt caacacggac
                                                                        780
  stctacaage acgtgttctg ggccacccgt ctggcgaaga agcttctcgg ctggttgctt
                                                                        840
  ttcaagaccc ccgatgaagg agcgtggact tccatctacg cagcagtcac cccagagctg
                                                                        900
  gaaggagttg gtggccgtta cctatacaac gagaaagaga ccaagtccct ccacgtcacc
                                                                        960
  tacaaccaga aactgcagca gcagctgtgg tctaagagtt gtgagatgac tggggtcctt
  gatgtgaccc tgtgatatcc tgtctcagga tagctgctgc cccaagaaac acattgcacc
                                                                       1020
                                                                      1080
 tgccaatagc ttgtgggtct gtgaagactg cggtgtttga gtttctcaca cccacctscc
                                                                      1140
 cacagggete tgteetetag ttttgagaca getgeeteaa eetetgeaga aetteaagaa
                                                                      1200
 gccaaataaa cattttggag gataatcacc ccaagtggtc ttcaaccata aactttgtga
                                                                      1260
 ttccaaagtg cccagttgtc acaggtgcca taaataatta cattttccaa cataaaaaaa
 1320
                                                                      1376
 <210> 67
 <211> 2434
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (10)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (12)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (27)
 <223> n equals a,t,g, or c
<220>
<221> SITE
<222> (73)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (75)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (103)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (130)
<223> n equals a,t,g, or c
<400> 67
etgggggtan tncaagaace etetgtngga ettagatgte aagetettte etttgggeag
                                                                     60
cgtgtttcct ttntncgagt agtgtgctgt gtaaactaaa ttngccggtt cgctttccat
                                                                     120
```

```
ttcctgacan ttgagatgga atgccttgac cattggtgct ctgacagaga agtcatggag
                                                                        180
 tcattgccat ttcctggttg cccttttgga atgtgatcct gttagtagag gttttctagc
                                                                        240
 ttctactaag atattcttt ccctaaccat catacacttg gcatgtttca ttcccatctc
                                                                       300
 ettteecete acettaaagg agactaecee tttgeeceat attgteaace taattteete
                                                                       360
 togtactoto totagtgaat gatgtgctac caagtatatg ccaggotgtg agaggattat
 actgagtagt agaaagaagc taatttgaaa taaaaattat ttgtataatt aagaaagcag
                                                                       420
                                                                       480
 attagatgca catggtcaac aggaagttga ctgtatgtct gctagttaga ttcaaaacat
 cataaagatg atagcatgtc aatatattag cctagccatt atgttagcct ttgttaggtg
                                                                       540
 ggcagctttt ctgctttttc ccttcctctg tggtgacaac ggaggaaata tccaacagaa
                                                                       600
                                                                       660
 atacgtctaa cagggaaatt gggatcatag tttatatgca tctgatttga aaggagtatt
                                                                       720
 gaggaaggtt ttcatatatg atctatcttt ggattaaaaa gaacatttat gaaatcaagc
                                                                       780
 cttctaacac tagttataat tgagaagcaa cagtaactcc gtggacagca atcaagctta
                                                                       840
 aaattgtaaa taaatatggg gataattcag ttgttgcaaa aaaagggcag aattcagtag
                                                                       900
 aataaagtcc ttttctctta caggtattaa atgaggacag agaacctcag gtgttcttat
                                                                       960
 gctagtgctt gctgagtgca tactaagaaa gcaattccaa atagatgtat acatctagag
                                                                      1020
agagtggtat tagagattca gtgtatgtat ttatttacat gagaggaaac tggaatataa
                                                                      1080
tcccataaat tattggaata taatcccata aattatcacc ttttatgact ggaaaatatt
                                                                      1140
tgccaatgaa gaaatggtct gtaggtattt gtcttaagat ttttggctgt ttaataaaaa
                                                                      1200
tgtaacttta acggtttctt atagttgcct ttataaagtg tattgtctaa aatatttttg
                                                                      1260
tatcatgtgc ctttgaaatt tgacagctga tttgggtgtt ggatttctgc ccagccattt
                                                                      1320
atcagtatta tcattttatt cagtagctgg caggtgtatt agacaaacga gacttaggta
                                                                      1380
aggaatggaa cettteetgt ggtttgaetg cacateacae cagaagaete cagtateeet
                                                                      1440
cattccagaa tgaggaaaaa gtattctaca aagaacctaa tcacctctgt gaaatctatg
                                                                      1500
ggatggaaac agtgtggcct taggagtcaa atagtctctg catggtgggg aggatcatga
                                                                      1560
tggaatatgt gaatttctac ttctagaagt tgtgaaatag gtcctgcact tttgcagaat
                                                                     1620
gteettettt aaacetgget tatteeacag etgtagetga taacatgace tggggettag
                                                                     1680
ctgctctagc cctgggttct tggagacctc acactgcctg gcccctggcc atccacctaa
                                                                     1740
ggactgcctg ctttctggtc acatgtggac cttgatacga ctaagcggtt acatatgtgg
                                                                     1800
ttgtgcaaaa gctttctgtt taatgcatag tgttaccgat ttacatcttg gttttcagtg
                                                                     1860
gcactatgtc taggaggcaa tatcctttta aacagtgctt tggctaagat agatacttgt
                                                                     1920
gaatcaaaga tagcacagaa atgaactaag tatatcccat ttggaattat attttgatac
                                                                     1980
tatttaaaat ggtttcacct gttaaagggc caacagaact cttggtttta cttttgtaat
                                                                     2040
tactgtacag aaaatttcaa gagtgtttga gtgcttgtca tcaggtgttt tccttaataa
                                                                     2100
gtagggatat gatcatttac aggaattata tatgaaaaaa gtttttgaaa tgtattttg
tgatgtgcta tgttgagggg aaaccaaata tttatgattt taaaacattc gtatgaaaac
                                                                     2160
attgtacaat gtaatatgct caactttctc aattttttgc taatttttct aagatacatt
                                                                     2220
                                                                     2280
aaaaatgttt tatatttttt tttaagtaaa atggacccag taagaaaatt aaaaatacca
                                                                     2349
2400
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa
                                                                     2434
```

```
<210> 68
<211> 1086
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (10)
<223> n equals a,t,g, or c

<220>
<221> SITE
<220>
<221> SITE
<220>
<221> SITE
<220>
<221> SITE
<222> (77)
<223> n equals a,t,g, or c
```

<222> (1056)

```
<223> n equals a,t,g, or c
 <400> 68
 ttgaaacacn cttttgagca ttaggtccca gctccaccgc ggtggcggcc gctctagaac
                                                                          60
 tagtggatcc cccgggntgc akkaattcgg cacgagcaac ctgtagtgag ttggctgtca
                                                                         120
 ctcagcagct ctggaaaatt acctgtctgc actgaatttc ctcatcagta aaatggaaat
                                                                         180
 gattatagta ctgaccttgt aggattattc taaaaatcag agaagttcat gcagcttaga
                                                                         240
 acagtgccag gcacatggca aatgctatgg catacttaag catcttcctc tgtggtgcct
                                                                         300
 cgtcatcacc atgtgattgt gccttgcttg tccctgtttc acttttcaga gggagaaaag
                                                                         360
 tggccaactt taagaatcaa aattctgatg ttacttcggg aaatgcatag agccagagag
                                                                         420
 acacaatttg acttagtatg atccacatca tcccctcagg ctgaatagtg gtggcatgca
                                                                         480
 catctatacc aaaatgtttt accttttttg tagaaggaaa atatttgtat cttctattcc
                                                                         540
 atatcttaga tctttataag agcacttaag ttcaacctcc taagaaactg ccaattttgt
                                                                         600
 tgatcatgat agtctgcaca gattttcgta ctatttagtg ktgggagtgc cttagggacc
                                                                         660
 atcaacaaca ggsccctcct tttatccatg agactactga ggccttggag gttatttgtc
                                                                         720
 catccatggc gggtgcacrg ctagaggtta tctggttagt agcccaacta agattagaac
                                                                         780
ccaggaattt tgatttaatc tcaaatgacc tcttttattt ctgcatccgg gaaggagaag
                                                                         840
 aggaaagtaa acgggaaatt catgttatct atggaaaagt tatcagtttg atgtttatta
                                                                         900
aatgatttgt ttcaggagat tgtttacaac ttttatcttc ctagacaata attttctgta
                                                                         960
agagtaaagt catggtcatt aaggtagtca tattaatgta ttcagtaasc tgtgaagaaa
                                                                        1020
aatatatatc aatgttttcc aataaaatac agtgantacc tgaaaaaaaa aaaaaaaaaa
                                                                       1080.
aaaaaa
                                                                        1086
<210> 69
<211> 1262
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (568)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (639)
<223> n equals a,t,g, or c
<400> 69
cctyctatag gkmaagctgg tackcctgca ggtaccggtc cggaattccc gggtcgaccc
                                                                         60
acgcgtccgc ttacaacttc tctyttgttt ctgtgtttct tccacaccat tatggcgtat
                                                                        120
ttactgatct acaaccggtt ttaagttagt tgcctgctta aaattttctc atgaattctc
                                                                        180
attacatttt caaacaaaat ccaaagtttt taccatgacc ttgtggatat ctttctgacc
                                                                        240
tcatttacta ctgttatctt ccttattcat tccattccac ccacagtggc cttcttgcaa
                                                                        300
gtcagagaat gcaacaaata tgctcctgcc ttggggcatt tgcacttctg ttcttttggc
                                                                        360
ctggtcactt cacctccaca ttctccattt tttatgactt tcttcctata tttggatctc
                                                                        420
tgttcaaatg tcaccctca aaaaggcctt caaaactacc ctatttaaaa tagcacctgc
                                                                        480
cactgccatc tccataccca ttttcatttt tccatagcac ttatccctac ttggcattat
                                                                        540
gttatatatt tgtttatttg kttactgnct gtctccccca ctggactgta agctccatga
                                                                        600
ctgtaagtac atagatgaat tacaaaatga atcaatggng aawtatcctt gmcatatatt
                                                                        660
atttaatgat tttgmctcca ttttaagtaa aaaaaaaata ccatttttt cactcttcaa
                                                                        720
agtgatatag tttaatcttc taaactacat ttttctcatt tcccgattta attaaatcag
                                                                        780
tgatataaaa aaaagagtga tggggatatg tgaaagaaga ctaaaataga tgccaggaaa
                                                                        840
tactaaaact gctgaagttt agtggtatat tttttcttta cacacagtat tatttgagtt
                                                                        900
actaattgtg tcactgaatt acagaataag caaaatacta ggtaaacaga atcacgcttg
                                                                        960
ggggctatat tttgtgtaaa atttgtgtta tgcaaaaata atattaaata tttaattact
                                                                       1020
acagttttgt tatttctttc ttattttagg aaatgatttg cagctgagtg aatcaggaag
                                                                       1080
tgacagtgat gactgaagaa atatytagct ataaataaaa atttatacag catgtataat
```

```
ttattttgta ttaacaataa aaattcctaa gactgaggga aatatgtctt aacttttgat
                                                                       1200
  gataaaagaa attaaatttg attcagaaat ttcaaaaaaa aaaaaaaaa aagggcggcc
                                                                       1260
                                                                       1262
  <210> 70
  <211> 1642
  <212> DNA
  <213> Homo sapiens
  <400> 70
 ggcgcgctgc ccggaretgc ctgggttgcg ctgccggcca cgtccccgcg ccgggcctca
 ggctccttcc tactgtccga gggccaccag gccgccgggg gcctgctgcg cccggatgcg
                                                                         60
                                                                        120
 tetgttacta gagtggagag tetacetteg tetcacatgt gecacaaagg atggcatgge
 cegggagtge cecaceaegt ggettteace ecetgeaaag ceagaetteg eceagegaea
                                                                        180
                                                                        240
 cagtgtcaag cccacagete tecaaggagg aagatggtee aggetgggag cateyeetta
                                                                       300
 gcagcagcct ctgatccctt ggccaagcag gagggaacca ttagcagcct gaggagctgg
 ctggctggga gcctcgggga ccgcccagcc ttgctcccag ctcacccaca agatgtggac
                                                                       360
 agetettgtg ctcatttgga ttttyteett gteettatet gaaageeatg eggeateeaa
                                                                       420
                                                                       480
 cgatccacgc aactttgtcc ctaacaaaat gtggaaggga ttagtcaaga ggaatgcatc
                                                                       540
 tgtggaaaca gttgataata aaacgtctga ggatgtaacc atggcagcag cttctcctgt
 cacattgacc aaagggattc ggcagcccam ctcaactcta tggaagtcac aacagaggac
                                                                       600
 acaagcagga cagatgtgag tgaaccagca acttcaggag gtgcagctga tggtgtgacc
                                                                       660
 tecattgete ccaeggetgt ggeetecagt acgaetgegg cetecattac gaetgeggee
                                                                       720
                                                                       780
 tocagtatga ctgtggcctc cagtgctccc acgactgcag cctccagtac aactgtggcc
                                                                       840
 tccattgctc ccacgactac agcctccagt atgactgcgg cctccagcac tcccatgaca
 cttgcactcc ccgcgcccac gtccacttyc acagggcgga ccccgtccac taccgccact
                                                                       900
 gggcatccat ctctcagcac agccctcgca caagtgccaa agagcagcgc gttgccaaga
                                                                       960
                                                                      1020
 acagcaaccc tggccacatt ggccacacgt gctcagactg tagcgaccac agcaaacaca
                                                                      1080
 agcagececa tgageacteg tecaagteet tecaageaca tgeecagtga caeegeggea
                                                                      1140
 agccctgtac cccctatgck tecccaagca caaggteeca ttagecaggt gteagtggac
                                                                      1200
 cageetgtgg ttaacacaac awataaatee acameeatge eetcaaacac aacemewgag
cccctcaccc aggccgtggt agacaaaact ctccttctgg tggtgctgtt actcggggtg
                                                                      1260
                                                                      1320
accettttca tcacagtett ggttttgttt geeetgeagg cetatgagag etacaagaag
aaggactaca cccaggtgga ctacttaatc aacgggatgt atgcggactc agaaatgtga
                                                                      1380
1440
gaccaaacca agtgcttcca aattcttttg gtgcaattga ggagatatgc cagatgctta
                                                                     1500
aacacattta attgctgtca gattaattcc atgatcacta aagagttgct gctttttca
                                                                     1560
                                                                     1620
taaaaaaaa aaaaaaaag gg
                                                                     1642
<210> 71
<211> 921
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (4)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (9)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (11)
```

```
<223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (15)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (20)
 <223> n equals a,t,q, or c
 <220> -
 <221> SITE
 <222> (901)
 <223> n equals a,t,g, or c
 <400> 71
gcgnggggna naggnaagcn ccccactatt gggttcaaaa gctggagctc caccgcggtg
gcggccgctc tagaactagt ggatcccccg ggctgcagga attcggcacg aggtctgagc
                                                                    120
agataagatt aagggctggg tctgtgctca attaactcct gtgggcacgg gggctgggaa
                                                                    180
gagcaaagtc agcggtgcct acagtcagca ccatgctggg cctgccgtgg aagggaggtc
                                                                    240
tgtcctgggc gctgctgctg cttctcttag gctcccagat cctgctgatc tatgcctggc
                                                                    300
atttccacga gcaaagggac tgtgatgaac acaatgtcat ggctcgttac ctccctgcca
                                                                    360
cagtggagtt tgctgtccac acattcaacc aacagagcaa ggactactat gcctacagac
                                                                    420
tggggcacat cttgaattcc tggaaggagc aggtggagtc caagactgta ttctcaatgg
                                                                    480
agctactgct ggggagaact aggtgtggga aatttgaaga cgacattgac aactgccatt
                                                                    540
tccaagaaag cacagagctg aacaatactt tcacctgctt cttcaccatc agcaccaggc
                                                                    600
cctggatgac tcagttcagc ctcctgaaca agacctgctt ggagggattc cactgagtga
                                                                    660
aacccactca caggettgte catgtgetge teccacatte egtggacate ageactacte
                                                                    720
tyctgaggac tcttcagtgg ctgagcagct ttggacttgt ttgttatcct attttgcatg.
                                                                    780
tgtttgagat ctcagatcag tgttttagaa aatccacaca tcttgagcct aatcatgtag
                                                                    840
900
nccggtaccc agggcggaag a
                                                                    921
<210> 72
<211> 906
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (34)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (833)
<223> n equals a,t,g, or c
<400> 72
ggaaattctc cctcactaat tggaacaaaa gctngagctc caccgcggtg gcggccgctc
                                                                    60
120
tgagcagagg acaggccggg agttttccgg gaacggagga agagcagtgg aggctgccag
                                                                   180
gatgaggetg ctgtgtggcc tgtggctgtg geteteettg ctgaaagtee tgcaggeeca
                                                                   240
gaccccaacc cccctgccac tcccgccccc gatgcagagc ttccaaggaa accagttcca
                                                                   300
gggggaatgg ttcgtcctgg gcctggcggg caacagcttc aggccggagc acagggcgct
                                                                   360
gctgaacgct ttcaccgcaa cttttgagct aagtgatgat ggccgctttg aggtgtggaa
                                                                   420
```

```
tgcgatgact cgaggccagc actgtgacac atggtcttat gtgctgatac cggcagccca
   gcctgggcag ttcactgtgg accacggtgt gggcaggagc tggttgctgc ctcccgggac
                                                                        480
  gctggaccag ttcatctgcc tgggcagagc tcarggcctc tcggatgaca acattgtctt
                                                                        540
  cccagatgtg actggargtg ccctggacct carcagcctg ccctgggtgg cagccccagc
                                                                        600
  ctgaccactc agacagccgc ggcccccaag gcctgactct tcttgtggga gggcgaggct
                                                                        660
  ggtcacccca ggccagcgtc tgttgaagga tgaagcagct cctgtccggc ccagccctgc
                                                                        720
  ctcacagetg tgcgagetet geceteetea geteteaaae etgaataaat genecaagee
                                                                        780
  cagaaaaaaa aaaaaaaaa aaaaaaaaa ctcgagggg ggcccggtac ccaattcgga
                                                                        840
                                                                        900
                                                                        906
  <210> 73
  <211> 680
  <212> DNA
  <213> Homo sapiens
  <220>
  <221> SITE
  <222> (7)
  <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (9)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (15)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (16)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (22)
<223> n equals a,t,g, or c
<400> 73
cactcantng aacannagct cnagctccac cgcggtggcg gccgctctag aactagtgga
teeccegge tgcaggaatt eggcacgaga tatttegetg gaccetagaa aagceaccac
                                                                      60
gacctgtggg ccatgatgct accccaatgg ctgctgctgc tgttccttct cttcttctt
                                                                     120
ctcttcctcc tcaccagggg ctcactttct ccaacaaaat acaacctttt ggagctcaag
                                                                     180
gagtettgea teeggaacea ggaetgegag aetggetget gecaaegtge teeagacaat
                                                                     240
tgcgagtcgc actgcgcgga gaaggggtcc gagggcagtc tgtgtcaaac gcaggtgttc
                                                                     300
tttggccagt atagagcgtg tccctgcctg cggaacctga cttgtatata ttcaaagaat
                                                                     360
gagaaatggc ttagcatcgc ctatggccgt tgtcagaaaa ttggaaggca gaagttggct
                                                                     420
aagaaaatgt tottotagtg ctccctcctt cttgctgsct cctcctycty cacctgctct
                                                                     480
cctccctacc cagagetetg tgktcaccct gttccccaga gcctccacca tgagtggagg
                                                                     540
600
aaactcgagg gggggcccgg
                                                                     660
<210> 74
<211> 1633
<212> DNA
```

```
<213> Homo sapiens
 <400> 74
 ggcacgagca tcagtaagta ggagctaggg aagagaaagc atgcaaaagg gcaagaatca
                                                                         60
 ggagaattga tatagtagct ggacagaact ctagatgtgt gtgtgtgaga gaaagagagg
                                                                        120
 cagggagaag gagggaggag ttacccccac gatgacctcc aacttccctt tctgcaccct
                                                                        180
 240
 gggctcaggg gtgggggagg ggctgcacca cattaggacc tgccgtactc caatcccatg
                                                                        300
 cagtcctcct gctcctgctg ctgcgtgcct gggctctggt catgccaggc ttccttgtgt
                                                                        360
 cctgcgtctg tggccggttc ctgccaacct gtctagtcct ttcaggcttg aggccctgca
                                                                        420
 ttgctctttc tggtcctctc ccctccttcc cgctcccat ctagcctttt ttgggttccg
                                                                        480
 ggatctgctg acagactttc ttcttgctgc ctgcctgctt acatttcaga agacccctct
                                                                        540
 ggaactgccc atggctgtgg tccacctgct ggtagcaacg ccctgttacc aaatgctaga
                                                                        600
 taatctgcca ctcccctctg cagccgccaa ctggtgctga gctgccacct gttgtctgtt
                                                                        660
 tccacaccct tgtacactgc attctgcctg tcactctgga cagctgccca gcagctgcag
                                                                        720
 ctgaggccgt cctgccagca cctaccgcag ccccatgggt ggtcctgtcc agttcctggc
                                                                        780
 gctcccagcc tgcctcatct gctgtactta cagagccaca tctggttgaa tacagagctg
                                                                        840
 gggtagcctg gatgggccca tggcacactc actcccagag aggcgggcat cctctctctg
                                                                        900
 gacctctagg gaagggcagg gtctggagcc caataataaa acacagtaat gataatcata
                                                                        960
 gctaatgttt actgagaact taggatgtga taggctcaat gatttaccag aattactcta
                                                                       1020
 tactgcataa ctaccctatc aggtaagtac tattaatatc ccctttccca catgaagaaa
                                                                       1080
 ctaaggctaa ccaatagaaa tgaacctgtg taatgtccca tgggcataat tcatggagcc
                                                                       1140
 aggattcaga accaggtgct ttacttcagg gctgaagctc ataaccacta gaccaaagcg
                                                                       1200
cetectecgt gacetetggg gaaggeecag caacateete tattgegtee aatgaettet
                                                                      1260
ccctggtctg agatccactg actcacaggg gtagggttaa ggttaaggcc agagtctcag
                                                                      1320
ctaagctttg gatactttct tctggatttg gagaaggctg gaataactga atttctgctc
                                                                      1380
atetteagga geggeetaet agageeacae attteecage teteettgtg tgttteecag
                                                                      1440
accettette categigies telecellag gelectggaa agiltleaga gagaalcace
                                                                      1500
cagtggtaac attgttaaac aaaacaggaa aatgggactt gtgtgtatat atgattaaat
                                                                      1560
tattaattga tggatctacc ttcttagctc gtgccgaatt cgatatcaag cttatcgata
                                                                      1620
ccgtcgacct cga
                                                                      1633
<210> 75
<211> 1022
<212> DNA
<213> Homo sapiens
<400> 75
ttcccgggtc gacccacgcg tccgcccacg cgtccggctt ggggccagca ccctgtctca
                                                                        60
aagatggcaa aatgaggcta gttctggatg agctagctgg tgtgggttcc aaccatagga
                                                                       120
acacactgat gctcaaatcc taaggtgcca agctctaggc cctggaggct ggtagaacag
                                                                       180
gatctatgcc tggaatcctg gcagggattc ctgtcaagga cttgtgttta agcctgcttc
                                                                       240
agggetteag getgettetg etetgtgtet geceaggetg getgageggg tggatgggtg
                                                                       300
gacagaaggg ctcaccaagg attgtggaca tagggtaggc cctggtacca cgggtttcag
                                                                       360
gctgttatca cttcccttgt aggaacatag ccagaagcag atgagccagg gtagagggct
                                                                       420
ggcccctcct ctcatcttcc cttcagtctt aaattgtctc cagcgatggg aagaggccag
                                                                       480
ggactgtaac cottgtgctg tgtattotot gagoototgo toactotoag ggocaagoag
                                                                      540
ctcccaagcc ggggccctct cttggccaaa atctgaggag cagtctaggt tacaggcttt
                                                                       600
ttggtaggta ggttctggct gcctgttaat gcagttaggc cccctgatta ggtacagtga
                                                                      660
gaaacaagct agaacaaccc tggcccagaa gactgtgcac tccagcaaga tccagggatg
                                                                      720
atageettge agggeeactg ggagtttgtg cecaagette tecetettet etececaggg
                                                                      780
ggcactggga ctggtccctg ccctcatcct tagcctgggc cttccccaga ggtattaaag
                                                                      840
agaagtatga ttcctctgtc ttcagttctt ttcaggggca tcctgcccat agtacccagt
```

teccaagggg ecceeagtea egtggtgaag ectageacte atgeagetet tagggaacea

900

960

```
<210> 76
   <211> 1184
   <212> DNA
   <213> Homo sapiens
  <400> 76
  agcaaaaggt gaagagaaa aggaagcttt ttctctaaaa atggtgcagc tatcctctga
  accaatttcc ttcggtttaa tgtacctgta tcttggggtt tttttccact taatttatcc
                                                                           60
  tggagctctt tccataacaa cacttggaaa gcactctcat cctttttca ctgctgaaca
                                                                          120
  gaattccact gtgtggatgg aacatactct atttcaccag tcccctgtag ccagtcactt
                                                                          180
  ggtttgtttc caatcttttg ctttttcaga gtaataacct tgtatgtcta tcattttgta
                                                                          240
  tgcatacagg tttatatgta ggaaaaattc ctagagtagg attgctggac caatggataa
                                                                          300
  aagtatattg tggacagaca atgccaaatt gcctttcaga gactgtggcc ctgtgcaccc
                                                                          360
  catcaggcat gtgtgactac caaagctect gtcagctgtt ttattttatc tcctttccag
                                                                          420
  totcaggotc aatgcagaac tttgaggtaa gottttctaa aatgtaggot cctaaacgcc
                                                                          480
  acagecaget etgecacatg aaggagaget caaatgagae agaaacagee tetgggcagg
                                                                          540
  atttctatcc tgcacagata tattttccac attctgggaa accgtgaagc ttccagagcc
                                                                          600
  acaattcccc agaaacacat ccccctgctg gtacagccaa gccccagaac aagctgtgct
                                                                         660
  tgcctggcac ctcaaagcca agcaccatgg atgccacttg ccatgggtgc ctgcaatttc
                                                                         720
  aaataatgag aaataagaaa tttcagcttc tcagtccctc tagccaacat ttcaggtgca
                                                                         780
 tgacagecte aggtggcaag cagetactet geeggacagg geagaagatg gaacacecea
                                                                         840
  tccctgggga agatcacgag gtcaggagat caagaccatc ctggctaaca cagtgaaacc
                                                                         900
 ccgtctctac tgaaaataca aaacattagc tgggtgtggt ggcaggcacc tgtagtccca
                                                                         960
 gctactcagg aggctgaggc aggagaatgg cgtgaacctg ggaggtggag gttgcagtga
                                                                        1020
 gccaagattg cgccactgca ctccagcctg gcaacagagt aagactccat atcaaaaaaa
                                                                        1080
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaa aaaa
                                                                        1140
                                                                        1184
 <210> 77
 <211> 312
 <212> DNA
 <213> Homo sapiens
 <400> 77
 aattcccggg tcgacccacg cgtccgtgat gagtggattt gtactcttac ccaggtcctg
 agggccagcc cacccagcat coccaccct gatgacgctg tecetacaac tggctgaact
                                                                         60
 ggtgcatttt gtgtgtgcct tccagagcca gtggactggt gtgtatccaa tgatgccacc
                                                                        120
 totgaaacct acagaaccac tatgetttge atgtgtacce tgcagggtet gagggccagg
                                                                        180
 ctgtctggta gctctgctcc tgggtgacag agcaagactc tgtctcaaaa aaaaaaaaa
                                                                        240
agggcggccg ct
                                                                        300
                                                                        312
<210> 78
<211> 1370
<212> DNA
<213> Homo sapiens
<400> 78
tggttaaaga gtacgacatt ttagtcaggg ctgagcacag cgcctggccg aggagtgctt
ttatatgctc ttggaagtat atggggactc catttctgtc actgtggcga ttccacttat.
                                                                         60
gcattcacca tgccaatcca aggcagctga tggcttagga aagtcagaga ctgagatgtt
                                                                        120
aaaatccttg gggctatcta ccaacatgtc tccattccac ctgttagggt taaaggtttt
                                                                        180
tctaacctgg gccctgacct tagcacagat ctgcctatat ttttttgaag ttcagccact
                                                                        240
tggactatta gccctaaact ttttctgtac tgccactgca gggctgaagg agctttgcat
                                                                        300
gcatccacca agtctggctt tcacacctga atttcacacc tcgctttcac ccttagctat
                                                                       360
tocatctttc tgtggaacat cagtgtcact tagcaatagc catacaatcc cattatcctt
                                                                       420
atacctacct ttcccttcaa agtctcggat gcctgataca ttgcacctgc tagtgcattc
                                                                       480
actoccatta gtacactocc aagtoottoc agtgaaagat gtaacaattg aatggccact
                                                                       540
gtgccaaagg tgcctgggct ctacctgcca ccagtgatgg ggtcctcaca gccaacccag
                                                                       600
                                                                       660
```

```
tggtgtatca tccatttcca aaaccatctc amtgcctgtt ttcttggacc actcctggca
                                                                         720
 tcaactgttt caggttagag tgactgaaaa tttgggktat aagatattta ttagagatca
                                                                         780
 atatctatta aaaatgtgaa aggaagcagg attktgctga ggaagaagat aacaacaaag
                                                                         840
 ataccacaat gtcagcccat caaagccttt ggccaaccca gcasaaaaat ctggagcagg
                                                                         900
 tgtayctttt tcagagtctc cccagttagg tcaaaatgtg cgcctttaca cccgcacttc
                                                                         960
 cctcaatcac gggcttcagg ctgtcctggg catgacctca gatgaagcgg ctcacacagc
                                                                        1020
 tgaggctaac gttgtcggag ctgacagctg aaagccgttt gctgaccaca ctcccacagc
                                                                        1080
 cgagcagcat gcccttgctt ggaggaggat ttggatgaca cagctctatg tctgccgtat
                                                                        1140
 tttgggggtc acatteetea cacceagete tgeeteaaat ggcaacattg acaaaaatte
                                                                        1200
 attccagttg caaaccaatt gtagaatact ataaaaccag agtacaagtc taatagcata
                                                                        1260
 aatctcatcc taagtggaaa agaaagggtc atttattcac acatttttgg ggaaaaaaaa
                                                                        1320
 acaacctttg ctatgtcttt attacaacac ggatactcac aaaaaatagt
                                                                        1370
 <210> 79
 <211> 368
 <212> DNA
 <213> Homo sapiens
<220>
<221> SITE
<222> (5)
 <223> n equals a,t,g, or c
<220>
<221> SITE
<222> (13)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (35)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (351)
<223> n equals a,t,g, or c.
<220>
<221> SITE
<222> (367)
<223> n equals a,t,g, or c
aatgnaactc tcnttactaa ttggaaccaa agctngagct ccaccgcggt ggcggccgct
                                                                         60
ctagaactag tggatccccc gggctgcagg aattcggcac gaggtccagg tgtctttatt
                                                                        120
tcagatgttc tgtttttcct ctattttctg ctcacatgaa catacacact tgccaggcac
                                                                        180
attetggett ttettattte tetttetgat tttgeeteet teetgeecet gttttettee
                                                                        240
cttttctctg gctatagaaa ctgtcaggtg gccttgctgg catcatccta cctcttttga
                                                                        300
actgtgttac ccaggaactt ccatttatta tgcctcgagg gggggcccgg nacccaattc
                                                                        360
ggaagant
                                                                        368
<210> 80
<211> 1088
<212> DNA
<213> Homo sapiens
```

```
<220>
   <221> SITE
   <222> (1)
   <223> n equals a,t,g, or c
   <220>
   <221> SITE
  <222> (4)
  <223> n equals a,t,g, or c
  <220>
  <221> SITE
  <222> (5)
  <223> n equals a,t,g, or c
  <220>
  <221> SITE
  <222> (9)
  <223> n equals a,t,g, or c
  <220>
  <221> SITE
  <222> (11)
  <223> n equals a,t,g, or c
  <400> 80
 nttnnaacng naaatctccc cttactattg ggaacaaaag ctggagctcc accgcggtgg
 eggeegetet agaactagtg gateeeegg getgeaggaa tteggeacga gatteeecag
                                                                          60
 aattggttca tcttgggaaa ggaaggccta ggaagctgat gacctacttg ttttgttcat
                                                                          120
 ccatttetet actgettett aaagtgeact caageggtea ccaggacate agaaaageca
                                                                         180
 aatccaaggt cccaaggctc ttgatcatcc agtgcccaca acagagggag taagctagac
                                                                         240
 ggagaagetg aagtgaagga ggaaagagag aagggggtge etgttecace ecetteetag
                                                                         300
 cagetgeete tgtgggetee ataateeeca tgeeetetee eecaceecae eeteeggeeg
                                                                         360
 mccacctete etecacceae eccatteate agtaggaggg caetgtatea geacageett
                                                                         420
 tcaaggcetg acagtteett tetgggtgee ecceacatgt aaccaettaa ecceccageae
                                                                         480
 aaagggccag cctgaatggg cttggagccg gtggctcgtg cgraggggga ggcctcaata
                                                                         540
 ggatttgggg agctgggtgt ataaaccgct ctgctccagc agccctcaat ggaggagctg
                                                                         600
 gagttggggc atgtgccgag gggaggagct cggcagctgc atcccgsccc tgaacaatga
                                                                         660
 tttattcatt ctcctggcac gcacagagaa gacaggctgg ggaaacatgg ccccacaaat
                                                                         720
 cgccctcagc ctcagctgat tcccacggca atcattgcag ttcagagtgc cccctccac
                                                                         780
 ttgtaatgtt ttgcctcaag cttgcccttc ctcttcctgg taaggaagta gactacatta
                                                                         840
 egggaggeet taagtageet taaatteagg caageategt tetttgagtt ggttgeacee
                                                                         900
aaactcaatc tcactctctc aaagctttct ggccaggttt tcctcacaca cccatagaaa
                                                                         960
cagatgcaga caaaaaaaaa aaaaaaaaaa actcgagggg gggcccggaa cccaattcgg
                                                                        1020
aagatttg
                                                                        1080
                                                                        1088
<210> 81
<211> 1862
<212> DNA
<213> Homo sapiens
<400> 81
ggcacgaget etgcaaaget eestetgete eagcacgtee tgtatttttt eaactttgga
tgtttgcaca agctattatc ttttgcaaga atgtcctgcc ccctcctatc catctgtgga
                                                                         60
agtcaacctt tetttaggge ceagaggaaa acetateace ttetacacet teccagatge
                                                                        120
ccaccatttg ggttaaactc tgcctcctcc aggtttgcca tggtcttttc cctctttaa
                                                                        180
agcactggtc ccagcctatg ccattgtgtg tcaccttggc tcctgtctcc tactggctgt
                                                                        240
agatetttgg gargggagga actgggggaat tttcateetg gggetatggg atagtgetta
                                                                        300
                                                                        360
```

===

1500

```
CCCCtaagta taaaataagc aagaatcctt gaattgaaaa gctccagtga cttqatagct
                                                                         420
 tccaacagcc cctaaagatg gggtagtcca ctgctcccta tataggaagg aagaggttca
                                                                         480
 cagteteatg gatttageag cettegtage eteaggatet treceaatgt ecceagttet
                                                                         540
 cctgctgggc cagctgtggt ctctgagaat gctcagttcc ctgggctagg ctggcttatg
                                                                         600
 gggccctccc tcctggacct gggtccatgt gaccaagtgg ccagccagcg agagagccct
                                                                         660
 gggtgatgag catctggaag tcagcttgct tgtctgccct aatcacagcc ctgggctcta
                                                                         720
 ccctcaagta ccctactcac atggtccaag agggaaccac tcttctgtca ttccctcccc
                                                                         780
 ttctgctcct catcctaccc acccagctgt tctgggcaag atcctgtggg catgatggaa
                                                                         840
 tectecettt tetetaceat ceatecettt gateacetaa tectrearet geeeteeaac
                                                                         900
 caaggtgttc tttctatacc aaaaaatctg atcaggccat cactcctctg tttaaaacca
                                                                         960
 ttcaacggct acccactgcc ctcaggataa aaactaaaat tcttaacagg atttacaaat
                                                                        1020
 tetteacetg gecettttga gataagagga tttataatet ttteaaagee aacaageate
                                                                        1080
ctgagacaat cttaattttt gcagctgtcc aagctattta agattgatga ggtctgcamt
                                                                        1140
aagttgttaa atggtgatta atttatctgg cctgaatggc akgtgaattc caaggggatc
                                                                        1200
atttttctct accttccttg ctgcacattg ggacacagca aggccaggac acccttytta
                                                                        1260
ggtgactgsg taaccaggtt cattaaggca gtcagcgcag agaagaagca gaaaatgggg
                                                                        1320
aggagttgtg taaatgaaat tgccgataat accaaaggaa ccccaagaga ggttcagtct
                                                                       1380
cccaacaatt gcagctttgc tgcaacatgg cagggtggga caggggtagg tttggatctg
                                                                       1440
ggctgctttg ccmtggaagt ggggttacag ggcagggaga ggcttaccat cactgagcat
                                                                       1500
ctttaattga ctttgtcatt ttastgagaa cttacaaatt gagctattct tcattgagtt
                                                                       1560
atatttggaa gcaggttcaa tacataatac agtcctccct ctagaatttc tgctttatgt
                                                                       1620
ccacgtgatg cagcttgttc atttctgaat tccttttaat gacaaaaaca tttggaagga
                                                                       1680
atgtaaagac tcaaaaaggg gatggcatgt aaactatagt tggaacaata tgtggccatt
                                                                       1740
taaaagaagt gttgtccagg aggtggaggt tgcagtgagc tgagattgca ccactgcact
                                                                       1800
ctagcctggg ccatagaatg agactccttc tcaaaaaaaa aaaagaaaaa aaaaactcg
                                                                       1860
                                                                       1862
<210> 82
<211> 1618
<212> DNA
<213> Homo sapiens
<400> 82
ccacgcgtcc gcaaggagcc agaggccatg cagtggctca gggtccgtga gtcgcctggg
                                                                         60
gaggccacag gacacagggt caccatgggg acagccgccc tgggtcccgt ctgggcagcg
                                                                        120
ctcctgctct ttctcctgat gtgtgagatc cctatggtgg agctcacctt tgacagagct
                                                                        180
gtggccagcg actgccaacg gtgctgtgac tctgaggacc ccctggatcc tgcccatgta
                                                                        240
tectcageet ettecteegg eegeeeceae geeetgeetg agateagace etacattaat
                                                                        300
atcaccatcc tgaagggtga caaaggggac ccaggcccaa tgggcctgcc agggtacatg
                                                                        360
ggcagggagg gtccccaagg ggagcctggc cctcagggca gcaagggtga caagggggag
                                                                        420
atgggcagcc ccggcgcccc gtgccagaag cgcttcttcg ccttctcagt gggccgcaag
                                                                        480
acggccctgc acagcggcga ggacttccag acgctgctct tcgaaagggt ctttgtgaac
                                                                        540
cttgatgggt gctttgacat ggcgaccggc cagtttgctg ctcccctgcg tggcatctac
                                                                        600
ttcttcagcc tcaatgtgca cagctggaat tacaaggaga cgtacgtgca cattatgcat
                                                                        660
aaccagaaag aggetgteat eetgtaegeg eageeeageg agegeageat catgeagage
                                                                        720
cagagtgtga tgctggacct ggcctacggg gaccgcgtct gggtgcggct cttcaagcgc
                                                                        780
cagegegaga aegecateta cageaaegae ttegacaeet acateaeett cageggecae
                                                                        840
ctcatcaagg ccgaggacga ctgagggcct ctgggccacc ctcccggctg gagagctcag
                                                                        900
ctgatacggc atcctgcgag aagacctgcc ctcctcactg ggatcccctt cctqcctcct
                                                                        960
cccagggete tgccagggee ttgctcagte cettecacea aagteatetg aactteegtt
                                                                       1020
teccagggee tecagetgee etcagaeact gatgtetgte eccaggtget etetgeeect
                                                                       1080
catgcccctc tcaccggccc agtgccccga ctctccaggc tttatcaagg tgctaaggcc
                                                                       1140
cgggtgggca gctcctcgtc tcagagccct cctccggcct ggtgctgcct ttacaaacac
                                                                       1200
ctgcaggaga agggccacgg aagccccagg ctttagagcc ctcagcaggt ctggggagct
                                                                       1260
agagcaaagg agggacctca ggccttccgt ttcttcttcc agggtggggt ggcctggtgt
                                                                       1320
teccetagee ttecaaacce aggtggeetg ceetteteee cagagggagg eggeeteege
                                                                      1380
ccattggtgc tcatgcagac tctggggctg aggtgccccg gggggtgatc tctggtgctc
```

acagtcgagg gagccgtggc tccatggcca gatgacggaa acagggtctg accaagtgcc

-

aggaagacct gtgctataaa ccaccctgcc tgatcctgcc cctgcctgac cccgccacgc

```
cctgccgtcc agcatgatta aagaatgctg tctcctcttg gaaaaaaaaa aaaaaaaa
                                                                         1618
  <210> 83
  <211> 2034
  <212> DNA
  <213> Homo sapiens
  <220>
  <221> SITE
  <222> (14)
  <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (382)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (1999)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (2027)
 <223> n equals a,t,g, or c
 <400> 83
 actcactaag ggancaaagc tggagctcca ccgcggtggc ggccgctcta gaactagtgg
                                                                          60
 atccccggg ctgcaggaat tcggcacgag atctaccggg agakggaytc tgaaagggcc
                                                                         120
coggecageg tecetgagae gecaacggea gteactgece cecattecag etectgggat
                                                                        180
acgtactate ageccegtge cetggagaaa catgetgaca geateetgge actggettea
gtattetggt ceatetetta ttaeteetet eeettegeet tettetaett gtacaggaaa
                                                                         240
ggttacttga gtttgtccaa agtggtgccg ttttctcact atgctgggac attgctgcta
                                                                        300
                                                                        360
cttctggcag gtgtggcctg cntccgaggc attggccgct ggaccaaccc ccagtaccgg
                                                                        420
cagttcatca ccatcttgga agcaacacat cggaaccagt cttcagaaaa caagaggcag
                                                                        480
cttgccaact acaactttga cttccggagc tggccagtcg acttccactg ggaagaaccc
                                                                        540
agcagccgga aggagteteg agggggeeet teeegeeggg gtgtggeeet gettegeeea
                                                                        600
gagcccctgc accgggggac agcagacacc ctcctcaacc gggttaagaa gctgccttgt
cagatcacca gctacctggt ggcgcacacc ctagggcgcc ggatgctgta tccaggctct
                                                                        660
gtgtacctgc tgcagaaggc cctcatgcct gtgctgctgc agggccaggc ccgactggtg
                                                                        720
                                                                        780
gaagagtgta atgggcgccg ggcaaagctg ctggcctgtg atggcaatga gattgacacc
                                                                        840
atgtttgtgg accggcgggg gacagctgag ccccagggac agaagctggt gatctgctgt
                                                                        900
gaggggaatg ctgggtttta tgaggtgggc tgcgtctcca cgcccctgga agctggatat
                                                                        960
tcagtcctgg gctggaatca tccaggcttt gctggaagca cgggggtgcc attcccgcag
                                                                       1020
aatgaggeta atgecatgga tgtggtggte cagtttgeca tecacegeet rggettecag
                                                                       1080
ccccaggaca tcatcatcta cgcctggtcc atcggcggct tcactgccac gtgggcagcc
atgtcctacc cagatgttag tgccatgatc ctggatgcct cctttgatga cctggtgccc
                                                                       1140
ttggccttga aggtcatgcc agacagctgg aggggcctgg tgaccaggac cgtgaggcag
                                                                       1200
                                                                       1260
catctcaatc taaacaacgc ggagcagctg tgcagatacc agggtcctgt actgctgatc
                                                                       1320
cggagaacca aggatgagat catcaccacc acggttcctg aggacatcat gtccaaccga
ggcaatgacc teetgetgaa geteetgeag categgtate eeegggtgat ggeagaggag
                                                                       1380
                                                                       1440
ggtcttcgag tggtgaggca gtggttggag gcctcctcac agctggagga agcctcaatt
                                                                      1500
tatagecgat gggaggtgga agaggaetgg tgtetgtetg teeteegete etaceaggea
gaacacggge cegaetteee etggagegtg ggggaggaea tgagtgeaga tggaeggegg
                                                                      1560
                                                                      1620
cagctggett tgtttetgge teggaageat etgeacaact ttgaggeeac teactgeace
                                                                      1680
ccacteceag eccagaactt ecagatgeee tggeacetet agggaceaac tgggacteat
                                                                      1740
```

gttcatgttg tcctcttgca tgcattacat	g ctgtttatag a cgtttcccct : gaatggatta	y tttgtggaaa : cattcatgtg	gtgggggaco gctgtactta ttaataaaaa	atoccctton accttctccan ggtattttt	gattetetgt teaceactgt a acatacatec ctacaaaaaa atce	1800 1860 1920 1980 2034
<210> 84						
<211> 2240)					
<212> DNA						
<213> Homo	sapiens					
<400> 84			•			
	atcacattaa					
tttccttcat	gcecgected	tcatgctgcc	cgctcgccta	ccgttcaggc	tgctgagcct	60
gaggaggtgc	ggattegete	ccacggcagc	gegeeatgge	ccccgggagc	cgctcctgga	120
ttatgaccc	gtagecyccc	cctccttcca	geactcateg	agtetgggae	gcgagettee	180
gttcccggag	tacatected	agggctttgg atccggagcc	agaaggegge	gacatgcagg	agegtttet	240
gctccagcaa	cacaccccgg	aggaggagcg	gcaacccacc	cycyaaaagc	agetgeagga	300
acagcaaaac	Ctacggggg	ggtcccggga	acagaggcag	atagagagaga	aggageggeg	360
attgccgccc	agcggcgtga	actgctcggg	ctacaaaaaa	grgggggacc	cggacccggc	420
cggagtgccc	ggctacctgc	cccgagagaa	attectecae	accoccacac	cagaggagg	480
gctggcacgg	accototocc	agcgctgctg	actactatea	caccaccage	cagacygegg	540 600
cctgcaggtg	agccgcgagc	agtacctgga	actaataaac	accacattac	gagaaccaacg	660
cccctccctg	gtgctctaca	tggtggacct	gctggacctg	cccaacaccc	tactaccaa	720
cttgcccgcg	ctggtgggcc	ccaagcagct	gatcgtgctg	ggaaacaaag	tggacctcct	780
gccccaggat	gctcctggct	accggcagag	gctgcgggag	cgactgtggg	aggactgtgc	840
ccgcgccggg	ctcctgctgg	cccctggcca	ccaaqqqcca	cadcdcccd	tcaaggacga	900
gccacaggac	ggggagaatc	cgaatccgcc	gaactggtcc	cgcacagtgg	tcagggacgt	960
gcggctgatc	agcgccaaga	ccggctatgg	agtggaagag	ttgatctctg	cccttcageg	1020
ctcctggcgc	taccgtgggg	acgtctactt	agtgggcgcc	accaacgccg	gcaaatccac	1080
tctctttaac	acgctcctgg	agtccgatta	ctgcactgcc	aagggctccg	acgccatcga	1140
cagagccacc	atctcccctt	ggccaggtac	tacattaaac	cttctgaagt	ttcctatttg	1200
caacccaact	ccttacagaa	tgtttaaaag	gcatcaaaga	cttaaaaaag	attcaactca	1260
agctgaagaa	gatcttagtg	agcaagaaca	aaatcagctt	aatgtcctca	aaaagcatgg	1320
ttatgtcgta	ggaagagttg	gaaggacatt	cttgtattca	gaagaacaga	aggataacat	1380
tccctttgag	tttgatgctg	attcacttgc	ctttgacatg	gaaaatgacc	ctgttatggg	1440
tacacacaaa	tccaccaaac	aagtagaatt	gactgcacaa	gatgtgaaag	atgcccactg	1500
gttttatgac	acccctggaa	ttacaaaaga	aaattgtatt	ttaaatcttc	taacagaaaa	1560
agaagtaaat	attgttttgc	caacacagtc	cattgttcca	agaacttttg	tgcttaaacc	1620
aggaatggtt	cegeeeegg	gtgctatagg	ccgcatagat	ttcctgcagg	gaaatcagtc	1680
ageregger	acagtcgtgg	cttccaacat	cctccctgtg	catatcacct	ccttggacag	1740
aaaagaagge	etgtateaga	agcatgcagg	tcatacgtta	ctccagattc	caatgggtgg	1800
actogogogo	totaggar	ttcctcctct	tgttgctgaa	gacattatgt	taaaagaagg	1860
aacacctaat	tttaaggag	tggccgacat	caagttttcc	tetgeaggtt	gggtttcagt	1920
gaccgtccgg	CCCCCtctct	gactgcatct	ccgaggctat	acacctgaag	gaacagtttt	1980
aagtgtggc	tataaaacce	tgccatatat agaagcctcc	ttaacttate	aaaggacagc	gcatcaagaa	2040
aggaaagata	aatgtatgag	accgacettg	ttcactccac	atattaact	yyaayaagaa	2100
aacaaaatac	attgaatttg	tattaaacat	ataacccats	acactdactg	cattgaacac	2160
ccttaaaaaa	aaaaaaaaa		acaacycata	aacaaagctc	ccatteetac	2220
						2240

<210> 85

<211> 1488

<212> DNA

<213> Homo sapiens

<400> 85

cgccaagtt	ccggagggag	agggtagaaa	ctggagggg	tagacctata	actcacggga	60
ctgagggtco	ttttctcccg	ctcccaggag	gaacgagaat	: gaatatgact	caageceggg	60
ttctggtggd	: tgcagtggtg	gaattaataa	ctatectact	ctacacgac	atccacaaga	120
ttgaggaggg	ccatctggct	gtgtactaca	gaaaaaaaa	tttactaact	agccccagtg	180
gaccaggcta	tcatatcatg	ttgcctttca	ttactacatt	Cagatetgto	cagacaacac	240
tacaaactga	tgaagttaaa	aatgtgcctt	gtggaacaag	taataaaata	atgatctata	300
ttgaccgaat	agaagtggtt	aatatottoo	ctccttatcc	antatttast	atcgtgagga	360
actatactgo	agattatgac	aagaccttaa	tetteaataa	aatccaccat	gaggagga	420
agttctgcag	tgcccacaca	cttcaggaag	tttacattga	attotttoat	caaatagatg	480
aaaacctgaa	gcaagctctg	cagaaagact	taaacctcat	acceccaat	Ctanatatata	540
aggctgtgcg	tgttacaaaa	cccaaaatcc	cagaagccat	aagaagaaat	tttgagttag	600
tggaggctga	gaagacaaaa	ctccttatag	ctgcacagaa	acaaaaaatt	ataannaa	660
aagctgagac	agagaggaaa	aaggcagtta	tagaagcaga	gaagattgca	Caagtagaa	720
aaattcggtt	tcagcagaaa	gtgatggaaa	aagaaactga	aaagcgcatt	tetessates	780
aagatgctgc	attcctggcc	cgagagaaag	Cgaaagcaga	tactaaatat	tatggtage	840
acaaatatgc	cacctcaaac	aagcacaagt	tgaccccgga	atatotoga	catgetgeac	900
accaggccat	tgcttctaac	agtaagatct	attttggcag	Caacatccct	aacatatta	960
tggactcctc	atgtgctttg	aaatattcag	atattaggac	togaagagaa	accatguing	1020
cctctaagga	ggctcttgaa	ccctctggag	agaacgtcat	CCAAAACAAA	ageteactec	1080
gttgatgcaa	gaggtggaaa	tgttctccat	atcaagatgt	adcccasada	gagagcacag	1140
aacaatcatt	atacggactc	ttcagattta	cagagaactt	acacttcatc	tattagrage	1200
ctcctgcgat	agtcctgggt	gctccactga	ttggaggata	gagccagctg	teteagagaga	1260
aaatggtctt	ttcagccaca	gtcttatcaa	gtatcctata	tatatteett	totalactes	1320
tactcatgaa	tgaggaaagt	ctgatgctaa	gatactgcct	gcactggaat	attanage	1380
aaatatataa	caagctgtgt	tttcctaagc	tgaaaaaaaa	aaaaaaaa	gccaaacacc	1440
		,	•			1488
	•					
<210> 86						
<211> 3174						
<212> DNA						
< 213 > Uama						

<213> Homo sapiens

<400> 86

geggaegetg grescaaaca ctaaggeetg ageggtgaea ategaggega gatgatggte 60 aacagggaat gcctcgtggg agaaaaaaga caattttatt ctcagcgctg attttgagat 120 gatgggcttg ggaaacgggc gtcgcagcat gaagtcgccg ccctcgtgc tggccgccct 180 ggtggcctgc atcatcgtct tgggcttcaa ctactggatt gcgagctccc ggagcgtgga 240 cetecagaca eggateatgg agetggaagg cagggteege aggegggetg cagagagagg 300 cgccgtggag ctgaagaaga acgagttcca gggagagctg gagaagcagc gggagcagct 360 tgacaaaatc cagtccagcc acaacttcca gctggagagc gtcaacaagc tgtaccagga 420 cgaaaaggcg gttttggtga ataacatcac cacaggtgag aggctcatcc gagtgctgca 480 agaccagtta aagaccctgc agaggaatta cggcaggctg cagcaggatg teetecagtt 540 tcagaagaac cagaccaacc tggagaggaa gttctcctac gacctgagcc agtgcatcaa 600 tcagatgaag gaggtgaagg aacagtgtga ggagcgaata gaagaggtca ccaaaaaggg 660 gaatgaagct gtagcttcca gagacctgag tgaaaacaac gaccagagac agcagctcca 720 ageceteagt gageeteage ecaggetgea ggeageagge etgecacaca cagaggtgee 780 acaagggaag ggaaacgtgc ttggtaacag caagtcccag acaccagccc ccagttccga 840 agtggttttg gattcaaaga gacaagttga gaaagaggaa accaatgaga tccaggtggt 900 gaatgaggag ceteagaggg acaggetgee geaggageea ggeegggage aggtggtgga 960 agacagacct gtaggtggaa gaggettegg gggageegga gaactgggee agaeeecaca 1020 ggtgcaggct gccctgtyag tgagccagga aaatccagag atggagggcc ctgagcgaga 1080 ccagcttgtc atccccgacg gacaggagga ggagcaggaa gctgccgggg aagggagaaa 1140 ccagcagaaa ctgagaggag aagatgacta caacatggat gaaaatgaag cagaatctga 1200 gacagacaag caagcagccc tggcagggaa tgacagaaac atagatgttt ttaatgttga 1260 agatcagaaa agagacacca taaatttact tgatcagcgt gaaaagcgga atcatacact 1320 ctgaattgaa ctggaatcac atatttcaca acagggccga agagatgact ataaaatgtt 1380 catgagggac tgaatactga aaactgtgaa atgtactaaa taaaatgtac atctgaagat 1440 1500 agcttttggg gggtactttg gaagtgtcta ataaggtgtc acaatttttg gtagtaggta 1560

```
tttcgtgaga agttcaacac caaaactgga acatagttct ccttcaagtg ttggcgacag
                                                                         1620
  cggggcttcc tgattctgga atataacttt gtgtaaatta acagccacct atagaagagt
                                                                         1680
  ccatctgctg tgaaggagag acagagaact ctgggttccg tcgtcctgtc cacgtgctgt
                                                                         1740
  accaagtgct ggtgccagcc tgttacctgt tctcactgaa aagtctggct aatgctcttg
                                                                         1800
  tgtagtcact tctgattctg acaatcaatc aatcaatggc ctagagcact gactgttaac
                                                                         1860
  acaaacgtca ctagcaaagt agcaacagct ttaagtctaa atacaaagct gttctgtgtg
                                                                         1920
  agaatttttt aaaaggctac ttgtataata accettgtea tttttaatgt acaaaacget
                                                                         1980
  attaagtggc ttagaatttg aacatttgtg gtctttattt actttgcttc gtgtgtgggc
                                                                         2040
  aaagcaacat cttccctaaa tatatattac caagaaaagc aagaagcaga ttaggttttt
                                                                         2100
  gacaaaacaa acaggccaaa agggggctga cctggagcag agcatggtga gaggcaaggc
                                                                         2160
 atgagagggc aagtttgttg tggacagatc tgtgcctact ttattactgg agtaaaagaa
                                                                         2220
  aacaaagttc attgatgtcg aaggatatat acagtgttag aaattaggac tgtttagaaa
                                                                        2280
  aacaggaata caatggttgt ttttatcata gtgtacacat ttagcttgtg gtaaatgact
                                                                        2340
  cacaaaactg attttaaaat caagttaatg tgaattttga aaattactac ttaatcctaa
                                                                        2400
  ttcacaataa caatggcatt aaggtttgac ttgagttggt tcttagtatt atttatggta
                                                                        2460
 aataggetet taccaettge aaataaetgg eeacateatt aatgaetgae tteecagtaa
                                                                        2520
 ggctctctaa ggggtaagta ggaggatcca caggatttga gatgctaagg ccccagagat
                                                                        2580
 cgtttgatcc aaccctctta ttttcagagg ggaaaatggg gcctagaagt tacagagcat
                                                                        2640
 ctagctggtg cgctggcacc cctggcctca cacagactcc cgagtagctg ggactacagg
                                                                        2700
 cacacagtca ctgaagcagg ccctgtttgc aattcacgtt gccacctcca acttaaacat
                                                                        2760
 tottcatatg tgatgtcctt agtcactaag gttaaacttt cccacccaga aaaggcaact
                                                                        2820
 tagataaaat cttagagtac tttcatactc ttctaagtcc tcttccagcc tcactttgag
                                                                        2880
. tecteettgg ggttgatagg aattitetet tgettietea ataaagtete tatteatete
                                                                        2940
 atgtttaatt tgtacgcata gaattgctga gaaataaaat gttctgttca acttaaaaaa
                                                                        3000
 cttgtcacag ctctgttttt ttgcttttcc ccttcatcct ctattagatc cttaaagaat
                                                                        3060
 ccaccctcaa actgtggact aaggatggca gtttgcctta aaccttctgt ggaatgagat
                                                                        3120
 gggctgtcag gacgtcttca acccacagta ccccgaaagc tttgtgctca ccac
                                                                        3174
 <210> 87
 <211> 2780
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (2760)
 <223> n equals a,t,g, or c
 <400> 87
ggcgccgcgg gcgggggtc ccggggccac gtggacgccg acggctctgg tctgcggcgt
                                                                         60
gtggacactg ccagcggctg cagccgactt ggaatgacct gggagacaaa tacaacagca
                                                                        120
tggaagakgc caaagtctat gtggctaaag tggactgcac ggcccactcc gacgtgtgct
                                                                        180
ccgcccaggg ggtgcgagga taccccacct taaagctttt caagccaggc caagaagctg
                                                                        240
tgaagtacca gggtcctcgg gacttccaga cactggaaaa ctggatgctg cagacactga
                                                                        300
acgaggagcc agtgacacca gagccggaag tggaaccgcc cagtgccccc gagctcaagc
                                                                        360
aagggctgta tgagctctca gcaagcaact ttgagctgca cgttgcacaa ggcgaccact
                                                                        420
ttatcaagtt cttcgctccg tggtgtggtc actgcaaagc cctggctcca acctgggagc
                                                                        480
agctggctct gggccttgaa cattccgaaa ctgtcaagat tggcaaggtt gattgtacac
                                                                        540
agcactatga actotgetee ggaaaccagg ttegtggeta teccaetett etetggttee
                                                                        600
gagatgggaa aaaggtggat cagtacaagg gaaagcggga tttggagtca ctgagggagt
                                                                        660
acgtggagtc gcagctgcag cgcacagaga ctggagcgac ggagaccgtc acgccctcag
                                                                        720
aggccccggt gctggcagct gagcccgagg ctgacaaggg cactgtgttg gcactcactg
                                                                        780
aaaataactt cgatgacacc attgcagaag gaataacctt catcaagttt tatgctccat
                                                                        840
ggtgtggtca ttgtaagact ctggctccta cttgggagga actctctaaa aaggaattcc
                                                                        900
ctggtctggc gggggtcaag atcgccgaag tagactgcac tgctgaacgg aatatctgca
                                                                        960
gcaagtattc ggtacgaggc taccccacgt tattgctttt ccgaggaggg aagaaagtca
                                                                       1020
gtgagcacag tggaggcaga gaccttgact cgttacaccg ctttgtcctg agccaagcga
                                                                       1080
aagacgaact ttaggaacac agttggaggt cacctctcct gcccagctcc cgcaccctgc
```

gtttaggag	it tcagtccca	c agaggccac	t gggttccca	T taataacta	- tana	
gaacatact	a agcgtgagg	t atottott	g tgtgtgtgtt	t tteessees	ccagaaagca	1200
cagattctt	t attaagtta	a officient	a gtaaatgtgt	- ccccaayee	acacactcta	1260
acattttca	g tggcgatat	a teceette	a ccttctcttg	aactcatyg	cactgtgtaa	
tttgagact	a aaatagogt	t decocooce	a coulded	g algadallic	catggtttcc	1380
tgagtgatt	t tootoaaac	e gagggaaat	g aaattgctgg	actatttgtg	gctcctgagt	1440
gaaaggtgg	c cttataaca	a daycacacc	c aaagcatagt	ttacctgcco	: acgagttctg	1500
tactoror	a atcoateac	y tallyalyi	t cctctgatct	taaggtcaca	gttgactcaa	1560
taccttcac	a accastast	z cyyaycaya	t tgaaatgcaa	aaacccacac	: ctctggaaga	1620
ttagccgtg	a tossecou	g gayettetg	t tgctgtgaat	acttetetea	gtgtgagagg	1680
ataacttta	t statesta	gulaceece	g accgtgcctg	i agtaagagaa	tgctgatgcc	1740
Cacagggg	a assistate	ctgtcaaat	c agttactgtt	caggggatco	ttctgtttct	1800
tattaasta	a aacacycec	Lagitecte	a tgttaacacg	aagccagagc	ccacatgaac	1860
atctcatta	a chachtage	aagggtagg	c atggaaaatt	ccacgaggct	cattctcagt	1920
caccette	a ctcattgaaa	gattccagt	t gtatttgtca	cctggggtga	caagaccaga	1980
caggetttt	c caggectggg	, tatccaggga	a ggctctgcag	ccctactaaa	gggccctaac	2040
cagag c c c c	a yayılıcıga	LCCCGCCCC	cagtagtect	tttagagget	toctatacte	2100
gguetget	c aaggaggtcg	accttctaat	: qtatqaaqaa	taggatagat	ttgatgtgaa	2160
gaccaaaya	c agalgicagi	gggctgctct	gaccctaata	tacacaacta	toocaoctot	2220
cgatgctag	L greetetaac	ccatgctgto	: cttqtqatta	aacacctcta	tetecettaa	2280
gaacaagca	. alacaggett	aagctctaac	, ataggtgttt	gtccttttac	Categageta	2340
ceccata	a Laaccacttt	gcatccaaca	ctcttcaccc	acctcccata	COCSSOCO	2400
rgrggataci	- Lyycccaaag	taactggtgg	r taggaatett	agaaacaaga	ccacttatac	2460
rgcccgcccg	g ayycagaaga	taacagcagc	: atctcgacca	acctctacct	taaannaaat	2520
CLLCALLAA	- cacgtatggt	tcacagataa	ttctttttt	aaaaaaaccc	aacctcctag	2580
agaagcacaa	i ctgtcaagag	tcttgtacac	acaacttcag	Ctttgcatca	coagtettet:	2640
acccaaga	i aatcaaagtg	gtacaatttg	'tttqtttaca	ctatgatact	ttctaaataa	2700
	. cccaaaaaaa	aaaaaaaaa	aaaaaaaaa	aaaaaggggg	adadaaaaa	2760
ggggccccc	ccccccaaa				22222233311	2780
*						2/60
<210> 88						
<211> 1061						
<212> DNA						
<213> Homo	sapiens.					
<400> 88						
aattcggcac	gagagaagga	aatacatcaa	aatgcccaca	ttggttatct	atagaggata	60
gaacgaaaga	rggctttatt	ctcttgtttg	ctcttattaa	agcaatcaga	taatacttet	60 120
cergracica	yayccctggc	tgcttcctqc	ctaacctctc	ctacaaacta	ctataassa	120
agaaaagcct	taaacggaaa	tgtgggagag	aaggttggat	tcactttcat	atctttccac	180
ggttgtgacc	cctcaagtcc	tggttgcctt	tgctgttctc	tattaccttc	aaacaccag	240
ctcgtcttta	tttcttttt	agttttgtcg	gggttggctt	gatagatgtt	aacayccag	300
agccagatgt	gtctagcctt	gtcttttgaa	tgcaagattt	agatataga	tactto	360
ttagtggaca	tcagagtcac	tagtcaggat	gaaagagttc	ttaaatttaa	tacttagetg	420
ttctggtaac	gtcatgtata	gtgacggcg	catgtctaac a	agatagaaaa	teecagaaa	480
tggggtggtc	totoaatcac	agtttgggag	acattgactt	taggragecag (gtaagtettt	540
tcactagata	atagagatat	aatacadadc	tttgaaagct q	ratatata	gttctgaat	600
gtggcaatgg	ggagggttga	aasaataact	gttgggcctg t	ggtgtettga	gacagagcc	660
gggcctgaac	tcaagcagag	acctcadaac	cgaaaggtgg t	cccccggtg a	gagttgaaa	720
CCCCacacac	gagtactctc	cacactaca	cyaaayytyy t	ygaaggatg (cagcaagagg	780
ggagaacttt	aagtccagg+	tttatass=	gggtctgaat a	cacgtggga g	gtgagag	840
tgtgttgagg	CCGGGGGGG	taaatastes	agtgacttag t	gcggccata t	cattagaaa	900
Caggaggatg	acttagaaca	eggeleatgt	ctgtaatccc a	igcactttga g	gaggctgagg	960
tctctacass	aaaaaaaaaa	ayyayittaa	aaccagcctg g	jacaacatag t	gagagcctg	1020
cacaaa						
	uudaadada (aaaaacccga	aaaaaaaccc a	1		1061

<210> 89 <211> 1342 <212> DNA

<213> Homo sapiens

```
<400> 89
 ccacgcgtcc gggcgcgcgg gtgaaaggcg cattgatgca gcctgcggcg gcctcggagc
                                                                        60
 gcggcggagc agacgctgac cacgttcctc tcctcggtct cctccgcctc cagctccgcg
                                                                       120
 ctgcccggca gccgggagcc atgcgacccc agggccccgc cgcctccccg cagcggctcc
                                                                       180
 geggeeteet getgeteetg etgetgeage tgeeegegee gtegagegee tetgagatee
                                                                       240
 ccaaggggaa gcaaaaggcg cagctccggc agagggaggt ggtggacctg tataatggaa
                                                                       300
 tgtgcttaca agggccagca ggagtgcctg gtcgagacgg gagccctggg gccaatggca
                                                                       360
 ttccgggtac acctgggatc ccaggtcggg atggattcaa aggagaaaag ggggaatgtc
                                                                       420
 tgagggaaag ctttgaggag tcctggacac ccaactacaa gcagtgttca tggagttcat
                                                                       480
 tgaattatgg catagatctt gggaaaattg cggagtgtac atttacaaag atgcgttcaa
                                                                       540
 atagtgctct aagagttttg ttcagtggct cacttcggct aaaatgcaga aatgcatgct
                                                                       600
gtgagcgttg gtatttcaca ttcaatggag ctgaatgttc aggacctctt cccattgaag
                                                                       660
ctataattta tttggaccaa ggaagccctg aaatgaattc aacaattaat attcatcgca
                                                                       720
cttcttctgt ggaaggactt tgtgaaggaa ttggtgctgg attagtggat gttgctatct
                                                                       780
gggttggcac ttgttcagat tacccaaaag gagatgcttc tactggatgg aattcagttt
                                                                       840
ctcgcatcat tattgaagaa ctaccaaaat aaatgcttta attttcattt gctacctctt
                                                                       900
tttttattat gccttggaat ggttcactta aatgacattt taaataagtt tatgtataca
                                                                       960
tctgaatgaa aagcaaagct aaatatgttt acagaccaaa gtgtgatttc acactgtttt
                                                                      1020
taaatctagc attattcatt ttgcttcaat caaaagtggt ttcaatattt tttttagttg
                                                                      1080
gttagaatac tttcttcata gtcacattct ctcaacctat aatttggaat attgttgtgg
                                                                      1140
tottttgttt tttctcttag tatagcattt ttaaaaaaaat ataaaagcta ccaatctttg
                                                                      1200
tacaatttgt aaatgttaag aattttttt atatctgtta aataaaaatt atttccaaaa
                                                                      1260
1320
aaaaaaaaa aaaaaaaaa aa
                                                                      1342
<210> 90
<211> 770
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (690)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (762)
<223> n equals a,t,g, or c
<400> 90
ggtagggtcg ggkcgaccyr cgcgtccggc tcttcctcag gcggcggcca tggcgggaca
                                                                       60
ggaggatccg gtgcagcggg agattcacca ggactgggct aaccgggagt acattgagat
                                                                      120
aatcaccagc agcatcaaga aaatcgcaga ctttctcaac tcgttcgata tgtcttgtcg
                                                                      180
ttcaagactt gcaacactaa acgagaaatt gacagccctt gaacggagaa tagagtacat
                                                                      240
tgaagetegg gtgacaaaag gtgagacaet caeetagaae agtgeegtge tgetgetggg
                                                                      300
aagttgettt acacaacaca ggccacatgg gaaaggcccc agcagccttc agctccttcc
                                                                      360
tttctcctta aagagcaaca gggcttattc ttgtttttct tttttcaaaa gtgtggcctt
                                                                      420
tgggctctgc catctggggt gtggtgtggt atgtgggaag aagttcagag gaaccgttgg
                                                                      480
aaacgacgtt aggcatttta ccttttcagt aacattttat acatctactt gtcaatgtat
                                                                      540
ttgagacatt cacagccaaa agcctgggac tctttgtgaa ggtcctcctc amctctatct
                                                                      600
ttctttctct ctctctcaaa ctttccttaa agttctcatt gcctttgcac tgcttctgtg
                                                                      660
aacagtettt gteteeteee caeetttggn ggaaagtgeg gggeagteet ggteaagaea
                                                                      720
ctcatgccct ggcaatgtgg cctgcagaga atgttgttgt anccaccagt
                                                                      770
```

<u>.</u>

```
<210> 91
  <211> 1570
  <212> DNA
  <213> Homo sapiens
  <400> 91
  gatggtttta ctgaagttgc taaaagttta cagaaaagga agtgcaggaa catttcacaa
                                                                         60
  atctacaatc tgtgagtatc acatcctgta tagctgtmaa cactggaata aggaagggct
                                                                         120
  gatgactttc agaagatgaa ggtaagtaga aaccgttgat gggactgaga aaccagagtt
                                                                         180
  aaaacctett tggagettet gaggaeteag etggaaceaa egggeaeagt tggeaacaee
                                                                        240
  atcatgacat cacaacctgt tcccaatgag accatcatag tgctcccatc aaatgtcatc
                                                                        300
  aacttctccc aagcagagaa acccgaaccc accaaccagg ggcaggatag cctgaagaaa
                                                                        360
  catctacacg cagaaatcaa agttattggg actatccaga tcttgtgtgg catgatggta
                                                                        420
  ttgagettgg ggateatttt ggeatetget teettetete caaattttae ecaagtgaet
                                                                        480
  tctacactgt tgaactctgc ttacccattc ataggaccct ttttttttat catctctggc
                                                                        540
 tetetateaa tegecacaga gaaaaggttr accaagettt tggtgcatag cageetggtt
 ggaagcatte tgagtgetet gtetgeeetg gtgggtttea ttateetgte tgteaaacag
                                                                        600
                                                                        660
 gccaccttaa atcctgcctc actgcagtgt gagttggaca aaaataatat accaacaaga
 agttatgttt cttactttta tcatgattca ctttatacca cggactgcta tacagccaaa
                                                                        720
                                                                        780
 gccagtctgg ctggawctct ctctctgatg ctgatttgca ctctgctgga attctgccta
                                                                        840
 getgtgetea etgetgtget geggtggaaa eaggettaet etgaetteee tgggagtgta
 cttttcctgc ctcacagtta cattggtaat tctggcatgt cctcaaaaat gactcatgac
                                                                        900
                                                                        960
 tgtggatatg aagaactatt gacttcttaa gaaaaaaggg agaaatatta atcagaaagt
                                                                       1020
 tgattcttat gataatatgg aaaagttaac cattatagaa aagcaaagct tgagtttcct
                                                                       1080
 aaatgtaagc ttttaaagta atgaacatta aaaaaaacca ttatttcact gtcatttaag
 atatgtgttc attggggatc tcttgatttg cctgacattg acttcagcaa aagcacgggg
                                                                       1140
 ctgtaaatta ccatttacta gattagccaa atagtctgaa tttccagaaa acaaggcaga
                                                                       1200
 atgatcattc ccagaaacat ttcccagaaa atgtttccca gaaaactaga cagmatgatc
                                                                       1260
                                                                       1320
 attcaatgga tcacagtgaa gcaaaggaca caacttttta ttgtacccct taattgtcaa
                                                                       1380
 caggagttaa ctgatttgtt gtggtgctca gactttttta tacaggtgct agtgtttat
                                                                       1440
 cctatgtatt ttaactcatt agtgcataaa ggcaagcccc atataatgaa gtctcagggt
                                                                       1500
 1560
 aaaaaaaaa
                                                                      1570
 <210> 92
 <211> 2950
 <212> DNA
 <213> Homo sapiens
 <400> 92
cccggctccc gcccgctccc agccgggccc cccagcggtc ggcgggacgg ctcccggctg
                                                                        60
cagtetgeec gecegeeceg egegggggee gagtegegaa gegegeetge gaceeggegt
                                                                       120
ccgggcgcgc tggagaggac gcgaggagcc atgaggcgcc agctgcgaag gtggcggcgc
                                                                       180
tgctgctcgg gctgctcttg gagtgcacag aagccaaaaa gcattgctgg tatttcgaag
                                                                       240
gactetatee aacetattat atatgeeget eetaegagga etgetgtgge teeaggtget
                                                                       300
gtgtgeggge cetetecata cagaggetgt ggtacttetg gtteettetg atgatgggeg
                                                                       360
tgettttetg etgeggagee ggettettea teeggaggeg eatgtaceee eegeegetga
togaggagee ageetteaat gtgteetaca ecaggeagee eccaaateee ggeeeaggag
                                                                       420
cccagcagcc ggggccgccc tattacacyg acccaggagg accgggggatg aaccctgtcg
                                                                       480
ggaattccat ggcaatggct ttccaggtcc cacccaactc accccagggg agtgtggcct
                                                                       540
                                                                       600
gcccgccccc tccagcctac tgcaacacgc ctccgccccc gtacgaacag gtagtgaagg
                                                                       660
ccaagtagtg gggtgcccac gtgcaagagg agagacagga gagggccttt ccctggcctt
tetgtetteg ttgatgttea ettecaggaa eggtetegtg ggetgetaag ggeagtteet
                                                                       720
                                                                      780
ctgatatcct cacagcaagc acagctctct ttcaggcttt ccatggagta caatatatga
                                                                       840
acteacactt tgteteetet gttgettetg tttetgaege atetgtgete teacatggta
                                                                      900
gtgtggtgac agtccccgag ggctgacgtc cttacggtgg cgtgaccaga tctacaggag
                                                                      960
agagactgag aggaagaagg cagtgctgga ggtgcaggtg gcatgtagag gggccaggcc
                                                                     1020
gagcatecea ggcaagcate ettetgeeeg ggtattaata ggaageeeca tgeegggegg
                                                                     1080
```

ctcagccgat gaagcagcag ccgactgagc tgagcccagc aggtcatctg ctccagcctg tcctctcgtc agccttcctc ttccagaagc tgttggagag acattcagga gagagcaagc	
tectered agenticete trecagaage tettegagag acatteagga gagagraage	114(
CCCTTGTCAT Gtttctgtct gtgttggtat a	1200
cccttgtcat gtttctgtct ctgttcatat cctaaagata gacttctcct gcaccgccag	1260
gadagggtag cacgegeage tercacegea gatggggeet agaateagge ttgettggag	1320
georgaeagt gatetgaeat ceactaagea aatttattta aatteatggg aaatcactto	1380
ctyceedaa ctyagacatt gcattityty agetettyyt etgattigga gaaaggactg	1440
ctatecatti titiggigig titatggaag tgcatqtaqa qcqtcctqcc ctffqaaatc	1500
agactgggtg tgtgtcttcc ctggacatca ctgcctctcc agggcattct caggcccggg	
ggtctccttc cctcaggcag ctccagtggt gggttctgaa gggtgctttc aaaacggggc	1560
acatctggct gggaagtcac atggactctt ccagggagag agaccagctg aggcgtctct	1620
ctctgaggtt gtgttgggtc taagcgggtg tgtgctgggc tccaaggagg aggagcttgc	1680
tgggaaaaga caggagaagt actgactcaa ctgcactgac catgttgtca taattagaat	1740
aaagaagaag tggtcggaaa tgcacattcc tggataggaa tcacagctca ccccaggatc	1800
tcacagtag tctcctgagt agttgacgg taggacagga tcacagtca ccccaggatc	1860
tcacaggtag tctcctgagt agttgacggc tagcggggag ctagttccgc cgcatagtta	1920
tagtgttgat gtgtgaacgc tgacctgtcc tgtgtgctaa gagctatgca gcttagctga	1980
ggcgcctaga ttactagatg tgctgtatca cggggaatga ggtgggggtg cttattttt	2040
aatgaactaa tcagagcctc ttgagaaatt gttactcatt gaactggagc atcaagacat	2100
ctcatggaag tggatacgga gtgatttggt gtccatgctt ttcactctga ggacatttaa	2160
teggagaacc teetggggaa ttttgtggga gacacttggg aacaaaacag acaccetggg	2220
aatgcagttg caagcacaga tgctgccacc agtgtctctg accaccctgg tgtgactgct	2280
guergerage gradette ceatgetgea ggeeteeate taaatgagae aacaaaggae	2340
datgettatt gittatatt aagataattg cgtgggtcca aacactcete tteeteeagg	2400
deductive type the attribute the state of th	2460
decigate gygrader gaataggeac ccaaaagtee gtgactaaat tregttate	2520
terrigatay cadattatgt taagagacag tgatqqctaq qqctcaacaa ttttqtattc	2580
detegered gradedaya grittitic cottoaactt gottagaatt groctactor	2640
guadgetgat detgeatatg gaagteeere tteggtgaea ttteetggee attettgtt	2700
courtingly galgginggi tgigcccact tectagagin agacagetee tagtatataa	2760
datteeeyya gegteegtgg tteagagtaa aettgaagea gatetgtgea tgetttteet	2820
objective type court of the cou	2880
Tytycadaat addadtdaat ttgggcaaaa aaaaaaaaaa aaaaaaaaa aaaaaaaaa	2940
aaaaaaaaag	2950
	2330
<210> 93	
<211> 1722	
<212> DNA	
<213> Homo sapiens	
•	
<400> 93	
COCACOACC BORGO COCACOACCA COCACOACA COCACOACCA COCACOACA COCACOACCA COCACOACCA COCACOACCA COCACOACCA COCACOACCA COCACOAC	
system agage agage getaggeet agageaceae taccectana tactacacaa	
ggcacgagcc agagcaggct gctaggcctg gggccaccac tgcccctggg tgctacaccc agtgtgctgg gtcactggga acttcctgaa gtggtgtcac ctgaactggg cccccaagga	60
agegigetyg gicactggga acttectgaa giggigicac cigaactggg coccaagga	120
tggggtgcgg gcagtaccgc aggaagagga gcagccctt tgaagattga gagctgcag	120 180
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggctctgtg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggccgg	120 180 240
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggctctgtg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggccggg gggccgggcc	120 180 240 300
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggctctgtg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggccggg gggccgggcc	120 180 240 300 360
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggcccgggcc cgggggacag aatccgcccc cgaaccttca aagagggtac cccccggcag gagctggcag accyaggagg tgcgacagac ccgcggggca aacggactgg ggccaagagc cgggagccg ggcgcaaaagg caccagggc cgccaagagc cgggagccg ggcgcaaaagg caccagggc cgccaagagc ggcgcaaaagg caccagggc cgcccaaggg gccgcaaaagg caccagggc cgccaagagc gcggaacag accgaggcc	120 180 240 300 360 420
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggctctgtg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggccggg gagctgcag accyaggagg tgcgacagac ccgcggggca aacgactgg gcccaagagc cgggagcgc ggcgcaaagg caccagggc cgccagggc cgccaagggc cgccaagggc cgccaagggc cgccaagggc cgccaagggc caccagggc cgccaaggc accyaggagcgc cgccaagggc cgccaaggc caccagggc cgccaaggc accgccttgg ggctctcggg tgcgcattagc attgggtccg caccattagc	120 180 240 300 360 420 480
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggctctgtg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggccggg gagctgcag accyaggagg tgcgacagac ccgcggggca aacgagggtac ccccggcag gagctgcag accyaggagg tgcgacagac ccgcggggca aacggactgg ggccaaagag caccagggc cgccagggc ggccaaagg caccagggc cgccagggc ggccaaaggc caccagggc cgccagggc acggcttgg ggcttctgcg ggccttcggg tgcgcgtctc gcctctagcc atggggtccg cagcgttgga gatcctgggc ctggtgctgt gcctgtggg ctggggggt ctggtgctg cggtgctgt cgctgtgga	120 180 240 300 360 420
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggccgggcc cgggggacag accyaggagg caccagggcc cgggagcgcg ggcgcaaagg caccagggcc cgggagcgcg ggcgcaaagg caccagggcc cgggagcgcg ggcctcggg tgcgcaaggc cgccagggcc ggcccaaggcc cgccagggcc cgggagcgc ggcccaaggcc cgccagggcc cgccagggcc cgccaaggcc cgccagggcc cgccagggcc caccaggcc cgccagggcc caccaggcc cgccagggcc ctggtgcgc ctggtgggcc ctggtgggcc ctggtgggcc ctggtgggcc ctggtgggcc ccacacacaccc gtgaccaccacacacacacacacacacacacacacacaca	120 180 240 300 360 420 480 540 600
tggggtgcgg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggactgtgcg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggccgggcc cgggggacag aatccgccc cgaaccttca aagagggtac cccccggcag gagctgcag accyaggagg tgcgacagac ccgcggggca aacggactgg ggccaaagag caccagggcc cgggagcgg ggccaaaggc cgccaagggc ggccttcggg tgcgcattc gcctctagcc atggggtccg cagcgttgga gatcctgggc ctggtgctgt gcctgtggg ctggtgcg caggtgccg ggcgcaagaccctg gcccatgtgg caggtgaccg ccttcctgga ccacaacatc gtgacggcc agaccacctg gaaggggctg tggatgtcg tggatgtga gagcacctg gcgaagggct ccacaaggcc gagacgacctg gaaggggct tggatgtga gagcacctg gagagggct ccacaaggcc gagagggct gagagggct tggatgtga gagcacctg gagagggct ccacaagaga gagaggggga agaccacctg gagagggctg tggatgtcg tggatgtga gagaggggggggga cacaagagggacagagggacaaggggaggggaaggaggag	120 180 240 300 360 420 480 540
tggggtgcgg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcag aggccgggc cgaggacag accacgaga accacgaga aggccgggcc cgggggacag aatccgccc cgaaccttca aagagggtac ccccgggag aatccgggc cggagcgg accacaggacgggaggaggacgggaggaggacgggggggg	120 180 240 300 360 420 480 540 600
tggggtgggg gcagtaccgc aggaagagga gcagcccctg tgaagattga gagctgccag aggccggggc cgggggacag acccacaggacggaggagggggggg	120 180 240 300 360 420 480 540 600 660
tggggtgggg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcag aggccgggc cgggggacag acccgcgccc cgaaccttca aagagggtacgg gcagcagggacgggac	120 180 240 300 360 420 480 540 600 660 720
tggggtggg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcag aggccgggc cgggggacag acccggggcc cgggggacag acccagagacgggggggg	120 180 240 300 360 420 480 540 660 720 780 840
tggggtgcgg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcaag aggacggcg gcagtaccgc cgagagaggag gcagccctg attggctgcg gcacgatgac ccgcgcacgg attggctgct tcgggggggggg	120 180 240 300 360 420 480 540 600 660 720 780 840 900
tggggtgcgg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcaag aggacgggg gcagtaccgc cgagagagagga gcagccctg ttggggggacag attggctgct tcgggggggggg	120 180 240 300 360 420 480 540 600 660 720 780 840 900 960
tggggtgcgg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgcaag aggacgggg gcagtaccgc cgaggacaga accacgagac ccgaggacaga accacgagac ccgaggacaga accacgagac ccgaggacaga accacagagac cacaagagacgagaggacgagaggacgagaggacgagaggacgagaggacgagaggacgagaggacgagaggacgagaggag	120 180 240 300 360 420 480 540 660 720 780 840 900 960 1020
tggggtgcgg gcagtaccgc aggaagagga gcagccctg tgaagattga gagctgccag aggactgcgg gcagtaccgc cgggggacag accyaggagg gcacgatgac ccgcgacggg attggctgct tcgggggggggg	120 180 240 300 360 420 480 540 600 660 720 780 840 900 960

ctcctgccag ccacgcctgc gaggcgttgg ataagcctgg ggagccccgc gcttccgccg ggtagcgcgg cgcgcaggct cctcggaacg tccggctctg	Caccccasca	1200 1260
cggctcctgg atccgctcct gcctgcgccc gcagctgacc ttctcctgcc	actagcccgg	1320
ccctgccctt aacagacgga atgaagtttc cttttctgtg cgcggcgctg	tttccatagg	1380
cagagegggt gtcagactga ggattteget teceetcaa gaegetgggg	gtcttggctg	1440
ctgccttact tcccagaggc tcctgctgac ttcggagggg cggatgcaga	gcccagggcc	1500
cccaccggaa gatgtgtaca gctggtcttt actccatcgg cagggcccga	gcccagggac	1560
cagtgacttg gcctggacct cccggtctca ctccagcatc tccccaggca	aggcttgtgg	1620
gcaccggagc ttgagagagg gcgggagtgg gaaggctaag aatctgctta	gtaaatggtt	1680
tgaactctca aaaaaaaaa aaaaaaaaaaa aa		1722
<210> 94		
<211> 635		
<212> DNA		
<213> Homo sapiens	٠	
223 Nome Suprems	•	
<400> 94		
gcctaaagag agctccccca ggaccagccc tggccaaggg attgctgcag c	recteateca	60
ccttccaagc actggaaaca aacattggag accaagtgag gcgtcactca a	cagccatag	120
- Lactragged datgardagt taratactga tatcettet tttgctgatt o	rgagt tagget	180
gearcyadda dydredgreg rgeecagrat traggagaag gaagegrett e	acctattat	240
tracetcage to be a cagetagged to the contract to the contract to the cagetage	atacctatt	300
acception of the second control of the secon	atgcagttt	360
difference according to agree and according agreement and agree according a	catecgeet	420
- coogggicea agegaticic cigicictae taaaaataca aaaattaget g	aatataata	480
grade tradiced ag clacicagaa ggctgaggtg caagaattgc t	tgaacctgg	540
adagegagg regerging cegagateae decacedeae tecadeetad d	caacagagc	600
tagactgtct caaaaaaaa aaaatgaccc tcgag		635
,		
<210> 95		
<211> 3798		
<212> DNA		
<213> Homo sapiens		
•		
<400> 95		
ccacgegtee ggggetteat acaggaaate tattgetgtg teaagtteea ga	anaaaannt	60
tctgttcgtc caagttacta accaggctaa accacataga cgtgaaggaa gg	Inactage	120
saddagagt geeceaetgt tgatggggta agaggateet gtactgagaa gt	traccara	180
suggetted coalgogoac agitectict graceagigt ggaggaaaag ta	actaaataa	240
wasacugada adyaydadac agaaatgctc tgcccttgga gaactgctaa co	tagggcta	300
- The state of the	Caaacaac	360
todetadege tycadactag caaggagaat catgetttag etteaageag th	tatotato	420
substitution of the state of th	cttcater	480
coogladaya oggotacaaa tgotgtqott tqttqccctc ctatcqcatt aa	maaatttm	540
accadada catgggaaat aatcetgaga ggccagcett cetgcacaaa ac	rectacase	600
uniqualcad algagaccaa ggaaaccaac tgtactgatg agagaataac ct	agatetee	660
agactegate agadttegga cetteagatt egtacegtgg ceatractea to	acadat a t	720
catagatyca taatggtaac acctgatggg aatttccatc gtggatatca cc	tccaagtg	780
stagetacae organizac congenticaa aacaggaata gaachgcagh ah	acaaaaca .	840
geographia agecageige geatatetee tagateeeag agggegatta ta	ccactaad	900
cadydatact gydgcaatgg cacagtgact gttaagagta catgccactg gg	addtccac	960
data de la contra del la contra della contra	tatacata	1020
sagetactic digiticiagg toccadada teageagaat taratatice at	atatoato	1080
colditatia tidittigac categigga ticattiggi tottoggangi ca	ataactac	1140
aguadatata datigaataa aacagaatet aeteeagtig tigaggagga tg	aaatooak	1200
decedegeea geedeacaga gaaqaacaat eetetetato atactacaaa ca	adatasaa	1260
gcatctgagg cattacaaag tgaagttgac acagacctcc atactttata ag	ttgttgga	1320

```
ctctagtacc aagaaacaac aacaaacgag atacattata attactgtct gattttctta
                                                                      1380
cagttctaga atgaagactt atattgaaat taggttttcc aaggttctta gaagacattt
                                                                      1440
taatggattc tcattcatac ccttgtataa ttggaatttt tgattcttag ctgctaccag
                                                                      1500
ctagttctct gaagaactga tgttattaca aagaaaatac atgcccatga ccaaatattc
                                                                      1560
                                                                      1620
aaattgtgca ggacagtaaa taatgaaaac caaatttcct caagaaataa ctgaagaagg
agcaagtgtg aacagtttct tgtgtatcct ttcagaatat tttaatgtac atatgacatg
                                                                      1680
tgtatatgcc tatggtatat gtgtcaattt atgtgtcccc ttacatatac atgcacatat
                                                                      1740
ctttgtcaag gcaccagtgg gaacaataca ctgcattact gttctataca tatgaaaacc
                                                                      1800
taataatata agtottagag atcattttat atcatgacaa gtagagotac ctcattottt
                                                                      1860
ttaatggtta tataaaattc cattgtatag ttatatcatt atttaattaa aaacaaccct
                                                                      1920
aatgatggat atttagattc ttttaagttt tgtttatttc ttttaagttt tgtttgtggt
                                                                      1980
ataaacaata ccacatagaa tgtttcttgt gcatatatct ctttgttttt gagtatatct
                                                                      2040
gtaggataac tttcttgagt ggaattgtca ggtcaaaggg tttgtgcatt ttactattga
                                                                      2100
tatatatgtt aaattgtgtc aaatatatat gtcaaattcc ctccaacatt gtttaaatgt
                                                                      2160
gcctttccct aaatttctat tttaataact gtactattcc tgcttctaca gttgccactt
                                                                      2220
tctcttttta atcaaccaga ttaaatatga tgtgagatta taataagaat tatactattt
                                                                      2280
aataaaaatg gatttatatt tttggtcatg tttgtaagag agtgaatgca cgtgtgagaa
                                                                      2340
cattagcttc ttctgaactc attatatctc cacagaggtg ttgatacttg atgcctaaca
                                                                      2400
gttttgcaga tgtgctacat tggaattgtg tatttttatg gtgtacattc tattgtgata
                                                                      2460
tatttattga ataattaatg totattgacc atataagtgg cgaaaaatgc accatagagg
                                                                      2520
acatggggta tttatttaca aactatgagc tacataataa gcaagtggcc atgggatggc
                                                                      2580
atgaccctcc cctccatatt tttgtggagc aaaatattgg caatgtttat gtaaatcatt
                                                                      2640
gttaatatca tgaaattatt tttaattaaa aacataagtc tatttgctcc atagcagaaa
                                                                      2700.
aaacatgaga agttttttca tcatgataga aattgaaaca aaatatattc attcttcaat
                                                                      2760
cataccatct gagattttta agacagctat tttgtcttat aagtatattt ttctccctct
                                                                      2820
agacatttca gttactatgg attttgtcct caaagggacc tttagtctat tttggatgta
                                                                      2880
aagctaatct aatgacactt ggcacatgat attttgatca agccattttg acttgaccaa
                                                                      2940
                                                                      3000
aaagcagtgt ccattaggtt tctgcatata aatattacca agcaatgttc acaatagaca
tcattacact gtccttgaaa tttattaatt cttcatccaa ccctggttga gctgaggctc
                                                                      3060
atagttaggt tcaagactat ctgtttaaat attactgaaa aacaaagtaa gacagtacta
                                                                      3120
tgcttacctc ttaacttgat aatgtcaaaa caggcatgtt aaatgacatc atagaaaaga
                                                                      3180
cttcaagata atttatagaa gttaaattat attgtacaga aaataattgt atgaaaatct
                                                                      3240
ctactatggg gctggaacat ggttgaacat tagaatgata taaaaaatta tatatattct
                                                                      3300
ccaaatccac gctagacctg tcaaattaga gaatctagag attagacctg gcgtgtcagc
                                                                      3360
aaggtcatcc aggaagcaga ggctgagacg gagttaggtg tgattactta catagtcgat
                                                                      3420
tacattttac aaataacatt ttatatgtct catttactgt gctttctccc catcccattt
                                                                      3480
tgtatetttt cetttgettt getagatttg teaattttet etetettet gtetetetet
                                                                      3540
ctttcaatat ctctaataat ttgaaagtaa ttcatcataa ctaaatatct attggggtta
                                                                      3600
                                                                      3660
tgcttcactt acaaacttct gaaaacggct ttactgagat ataattgata tatttaagtg
tacagtttgt taaattttgc acatatttaa aatgtggact ttggtaaatg ttgacatagt
                                                                      3720
tttacatctg tgaaaccatc agcataatca agataataaa cttgtccatc acccccaaa
                                                                      3780
                                                                      3798
aaaaaaaaa aaaaaaaa
```

```
<210> 96
<211> 2683
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (2640)
<223> n equals a,t,g, or c
<220>
<221> SITE
<220>
<221> (2640)
```

<223> n equals a,t,g, or c

```
<400> 96
   acaasgtmac gcctgacagg tmaccggatc cgggaattcc cgggtcgacc cacggcgtcc
   gcatttgcaa taacagaaaa ggaattgcat gtatgaagtt ttcaatcgtg ggcttttctt
                                                                            60
   tgttgtgggg aggggtcgg gggatagttt gatttccatt ttctgaaaac gacagacttg ·
                                                                           120
  gattetgttt gtgtgtgcat attttateca geettaagtt ataaagetea tetgteeege
                                                                           180
  tgcattccct gtgtattttc aggacatggc tcgtgggtgt gtgtgttcat tgtgtgcgtc
                                                                           240
  tgtatgtatt tttctgtcat cactgttccc tctcctcccg agtgtgcatt cagttaatat
                                                                           300
  aatcagttgc ttgcttcttt caaagtgctt tgaaggtctt gaactcatgt gtgagcatct
                                                                           360
  ttatcaacta tcccaattgc atgttctcca tcacatattc tcttatttgc tctgtacccc
                                                                           420
  ctgagaatat gttttagaga tattggaata aagctgtctg ggtaaggagt aggcttagcc
                                                                          480
  gacctatgaa taatacactt tagtctagtt ctttattcta aatctggatt gccagtattg
                                                                          540
  tgtatttaaa ccaagtctgt gaatacctgc tttttttggc cacagagtaa caagttttca
                                                                          600
  tgtaagatet teataceaaa gtaggaagta aaaatagett agaaagetet gteaggtgtt
                                                                          660
  ttgtgcagct gacagargta atgttacatc acctaaaaaa gaaagataca cggtcagtta
                                                                          720
  tcctaaaaat aaattgtttg gaaagtacaa tgcaccacat ttttgtagaa gtctactatt
                                                                          780
  tgataaacag ttgaaattca agatgtgttt gacccttagt catttttact ctttggttct
                                                                          840
  gagtatacct attttcttag cgtatctgcc ttgtttatct ttttcttcac cttttaacaa
                                                                          900
  gtatgacata ggaaagtcat tttttttag aattcatgga tcagtctgat ctactcttat
                                                                          960
  tcataatgga acatgtaaat atactgaaaa ctgtttttca ggagagaaat atgagttgga
                                                                         1020
  gggaaggaaa agtggttcta ctaatgttcc aaaatcctca tcagagaagg tatgatgttc
                                                                         1080
  tcaggtgtgg aaaatatttt ttagttgatt gagaatgcag gtttaacaga agagataagg
                                                                        1140
 ggcataatga ctgctggttt tccagactgg attttcctac cgcaactatt aatgttctca
                                                                        1200
 gagttgatga ggaccacctt tgtgtataca cttgtagttt taaaccttgc attggtaaca
                                                                        1260
 aaatgatcaa ctttaatcca ggtagaattc aagatggctg tacttcagtt gtatgataaa
                                                                        1320
 attaatggtt ctcatgactt gtgtggcatc taaaaataat gtttttatag catctctctg
                                                                        1380
 ccactaaatt gttgacttga attttgggaa aaaaaaaagt tggtgttgat atgtatatgt
                                                                        1440
 gtgtgtgtat atatgtattt ataaacaagt gtgtttgagt aacaagtgag tttcatagtc
                                                                        1500
 ttecectacg catgtgtatt ccacacacaa atggetgagt tatagteata aaacaatttg
                                                                        1560
 caataaaaaa aaaaccaaaa cagattgtca gttaaccagg aaacagttaa tgttttttaa
                                                                        1620
 tgaatctggc attatagtga gcaaatgtcg tattaattta ggctaatttc taatactacc
                                                                        1680
 ataatttgtg totaaattto tgttggggta gaaattacta aaattgtggg gagtttttto
                                                                       1740
 tgatttttac attgctttag gaaacatttt tactaattca gctgtcttag gtaaaatgaa
                                                                       1800
 tagttttctt cctgtttttt tatgtgtcat tgttagtggt ctcagaattc tgatcagtaa
                                                                       1860
ctttgtgtat gatgctgaat tacaaaccgt ttgaatgatc cagttgaaaa cgtatccctc
                                                                       1920
tactttcttc agttgtagaa aaggttaatt tccctcagtg tcccacatta taccaaccta
                                                                       1980
agagaagaac aggtaatagg gagaaataaa catacggtgg tttcagtggt tttggtcatg
                                                                       2040
tgtccacagg agaaactaac cattcagttg tcttaatttt agttcgttct accctgtgag
                                                                       2100
gagtttgttt ccatcagttg ttgactttcc aaaatgttgc attaagtaat agttgtcact
                                                                       2160
ctgttggtct catggtcaat atcaatcaga ctttcatgat ctctactaat tattagtaga
                                                                       2220
gtcctgtact atgtctgtaa ctactaagtt taaagaaaag cacatagtca cttcatctct
                                                                       2280
ttttttctta gcctacgctc actecccaac ccateccaac attgacatgc tatetgtgga
                                                                       2340
caaatagcag ttctcagaat ctagtcaagt tgccatcatc ccccttgcct tggccgttca
                                                                       2400
tagtaggtat gcatatgttt gtttctgtac agtactgtgt gtgtgtgtgt atatatatat
                                                                       2460
acatetgtat gcacacatet ttgataaaat agetatttga etageagggt taaagtgget
                                                                      2520
tttaattact tcgtgagtgt tattggatac atcttaaaaa aaaaaaatct ggacccagan
                                                                      2580
ccatgccata cttggttgga ctattttggg gcattnaaaa ttg
                                                                      2640
                                                                      2683
```

```
<210> 97
<211> 2181
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (5)
<223> n equals a,t,g, or c
```

<400> 97

gtacnggatt cccgggtcga cccacgcrtc cggcaggcag aggggcagtg ggcgcgggga agatgattct gggcctcccc catcactgt cattagggs and	
agatgattet gggcetecce catetactgt cattaaccaa aatgaaacat ttgccaacat	60
aatttttaaa cctactgtag tacaacaagc caggattgcc cagaatggaa ttttgggaga ctttatcatt agatatgacg tcaatagaga acagaggatt ganaa ttttgggaga	120
ctttatcatt agatatgacg tcaatagaga acagagcatt ggggacatcc aggttctaaa tggctatttt gtgcactact ttgctcctaa agaccttcct agatatga	180
tggctatttt gtgcactact ttgctcctaa agaccttcct cctttaccca agattctaaa attcgtgctt gacagcagtg cttctatggt gggaacgaaa attcgtgctt	240
attcgtgctt gacagcagtg cttctatggt gggaaccaaa ctccggcaga ccaaggatgc	300
cotottcaca attotccatg acctocgace coaggacegt tteagtatea ttggattte	360
caaccggatc aaagtatgga aggaccactt gatatcagtc actccagaca gcatcaggga tgggaaagtg tacattcacc atatgtcacc cactggagga aggacagt acatcaggga	420
tgggaaagtg tacattcacc atatgtcacc cactggaggc acagacatca acggggtcct gcagagggcc atcaggctcc tcaacaagta cgtggcgag acagacatca acggggtcct	480
gcagagggcc atcaggctcc tcaacaagta cgtggcccac agtggcattg gagaccggag	540
cgtgtccctc atcgtcttcc tgacggatgg gaagcccac ggtggcattg gagaccggag caagatcctc aacaacaccc gagaggcgg ccgagggaga gcacaccct	600
caagateete aacaacacee gagaggeege eegaggeeaa gtetgeatet teaecattgg	660
categgeaac gaegtggaet teaggetget ggagaaactg tegetggaga aetgeggeet caeaeggege gtgeaegagg aggaggaege aggetggaga attach and an	720
cacacggcgc gtgcacgagg aggaggacgc aggctcgcag ctcatcgggt tctacgatga aatcaggacc ccgctcctct ctgacatcgg catcgattat accounty	780
aatcaggacc cogeteetet etgacateeg categattat coccecaget cagtggtgca ggecaccaag accetgttee ceactactt caacggeteg good cagtgeteg cagtggtgca	840
ggccaccaag accetgttee ccaactactt caacggeteg gagateatea ttgeggggaa getggtggae aggaagetgg atcacctgea cgtggaggta	900
gctggtggac aggaagctgg atcacctgca cgtggaggtc accgccagca acagtaagaa attcatcatc ctgaagacag atgtgcctgt gcggggtaa	960 1020
attcatcatc ctgaagacag atgtgcctgt gcggcctcag aaggcaggga aagatgtcac aggaagcccc aggcctggag gcgatggaga ggggacmag aaggcaggga aagatgtcac	1020
aggaagcccc aggcctggag gcgatggaga gggggacmcc aaccacatcg agcgtctctg	1140
gagctacete accacaaagg agetgetgag etectggetg caaagtgaeg atgaacegga gaaggagegg etgeggeage gggeecagge etgeggtgtg aaagtgaeg atgaacegga	1200
gaaggagcgg ctgcggcagc gggcccaggc cctggctgt agctaccgct tcctcactcc ctcactcc atgaagctga gggggccggt cccacgcatg gaagatgacg atgaaccgga	1260
cttcacctcc atgaagctga gggggccggt cccacgcatg gacggcctgg aggaggccca cggcatgtcg gctgccatgg gacgcgaacc ggtggtgcag aggaggccca	1320
cggcatgtcg gctgccatgg gacccgaacc ggtggtgcag agcgtgcgag gagctggcac gcagccagga cctttgctca agaagccata ccagccaaga attaca	1380
gcagccagga cctttgctca agaagccata ccagccaaga attaaaatct ctaaaacatc	1440
agtggatggt gatccccact ttgttgtgga tttccccctg agcagactca ccgtgtgctt	1500
tgtcacagtg aacggagagt taattaana suggetggte tetgatcaca gggactetgg	1560
gcgcacttac ttgcgcacta tcaccataat actegecct ccaaatggcc acaagaaaca	1620
gatcacaccg agcagagtca tottogates tottogates	1680
gagtgtggtg gtggggaget gaggketes saysgattaga ttggtgttcc cctgcaacca	1740
Cgtcaccatc cagggttca taggettta agggttcagtg tctgccaacg ccaatgtcac	1800
gcccttccag cgacaccacc taggettata	1860
Ctgccacgga ctgctgggtc agttgctact actgctade agcgagggcc tttccagcaa	1920
agggcccagc cagaacctca ctcaccatat todayacycc agactcacag aagaccctgc	1980
cgtcctaaca gtgaaaggcc accaagtasa gttccttcag gtgggagagg ggcctgaggc	2040
cggggaagag cagwtagayt gytggtttgc caggaacatg ccgccaaact gattgacggg gagtacagga ttacctggca t	2100
gagtacagga ttacctggca t	2160
	2181
23.0	
<210> 98	
<211> 1957	
<212> DNA	
<213> Homo sapiens	
<400> 98	
acgcgtccgg aggcggcagg ctcatggcgc cggcgtcgcg gttgctcgcg ctctgggcgc tggcggctgt ggctctaccc ggctccgggg cggaggggg	
tggcgggggg cgtggcggag gaggagggt ggaggggga cggcggggtgg cgccgggggggggg	60
ggccgggggc cgtggcggag gaggagcgct gcacggtgga gggtcgggcg gacctcacct acgcggagtt cgtgcagcag tacgccttcg tcaggccggt catcgggcc gacctcacct	120
acgeggagtt cgtgcagcag tacgecttcg tcaggccgt catectgcag ggactcacct acaactcgag gttccgggcc ctgtgctccc gcgacaggtt catectgcag ggactcacgg	180
acaactegag gtteegggee etgtgeteee gegacaggtt getggetteg tttggggaca gagtggteeg getgageace gecaacacet actestagg	240
gagtggtccg gctgagcacc gccaacacct actcctacca caaagtggac ttgcccttcc	300
aggagtatgt ggagcagctg ctgcaccccc aggaccccac ctccctgggc aatgacaccc tgtacttctt cggggacaac aacttcaccg aggaggagta tatattat	360
tgtacttett eggggacaac aactteaceg agtgggeete tetetttegg cactacteec caceccatt tggeetgetg ggaacegete cagettacag obtatacteec	420
caccccatt tggcctgctg ggaaccgctc cagcttacag ctttggaatc gcaggagctg gctcgggggt gcccttccac tggcatggac ccgggtacts	480
gctcgggggt gcccttccac tggcatggac ccgggtactc agaagtgatc tacggtcgta agcgctggtt cctttaccca cctgagaaga cgcgagagtt gcccttaccca cctgagaaga cgcgagagtt gcccatgatca tacggtcgta	540
agegetggtt cetttaceca cetgagaaga egecagagtt ceaceceaac aagaceaege tggeetgget cegggacaca tacecageet geacegtetg	600
tggcctggct ccgggacaca tacccagct gcaccgtctg cacggcctg gagtgtacca tccgggctgg tgaggtgctg acttccggac cgctggtaga about	660
teegggetgg tgaggtgetg actteegget geacegtetg caeggeeetg gagtgtacea accagegtet teateteace treetegget aggearage getgetget caacettgae	720
accagegtet teateteace tteetegget agecaaaaca getggeagga etgeeggtea	780
	840

```
cacaccagca cgtcccacct cgtgctcacg gattttatta cacagatagt ggcggcaatg
                                                                       900
geoteagece ageocaccet cacetgettt tecageceae aaagggggae gateaeggee
                                                                       960
                                                                      1020
cagcaaaagc gatgctgaga ggggaaacag tccagagtcc aacagcagaa cttggggggaa
gcggtcgggg tggccaggaa cataaactat gtataggggc cgggggcttc tgcccagggc
                                                                      1080
tcccctggac caggacgcca ggtagggcag ggaacctcag tagtcctcca cccagccatt
                                                                      1140
ctcagagatg aatgcgtcaa taacctcctt catagccaag ttggggatga gctgttcctg
                                                                      1200
ggtcaggggg ctccgggtca cggggtcaaa atgacccaca cgctgcagtg acaagaaggg
                                                                      1260
cagagggcag tcatggggcc caggaccatg ccactggccc tgctccccca gccgcaggcc
                                                                      1320
teacctgeag gtgctccteg atgtccttgc ggtcgtaggt gatgccactg ggcgtgatgc
                                                                      1380
acggctcccg catcagctca aagctgatct tgccacacag gtagtcgggg atgtctcgct
                                                                      1440
tctgtggcac aggggcacac ggtcagaggc tgaaaagggg cactgcacga gcacctgcca
                                                                     1500
gccatcggca gcaagcgaca cacactcacc ttcctcttct catccacctg agaaaaaagc
                                                                     1560
1620
accecacete cetectecat ggggcacaga cecaacacaa ggcggggatg eteccaegee
                                                                     1680
acgtgcacac acacagaccc acatgtgggt ggggggcacc ctcacgtgct tggcctcaat
                                                                     1740
gcaggcctgc tgggcccgga cgtggctgtc gtcctcatca ccctcgtggt ttcgctggca
                                                                     1800
ctcttccagc tecetggggg ttgaccagga qeeggteaga gatqqaeetg geeagatgte
                                                                     1860
tgaccacacc ccaatctcag agctaacatc cacacttccc cacatttcct gcttgccagt
                                                                     1920
aaagccttcg ataaacaaaa aaaaaaaa aaaaaaa
                                                                     1957
<210> 99
<211> 1112
<212> DNA
<213> Homo sapiens
<400> 99
ccacgcgtcc gcccaccaac aaaggcactt gcagttacat tcaccacatt tgtaacggag
                                                                       60
ccattgaagc atattggaaa aggaactggg gaatttatta aagcactcat gaaggaaatt
                                                                      120
ccagegetge tteatettee agtgetgata attatggeat tagecateet gagtttetge
                                                                      180
tatggtgctg gaaaatcagt tcatgtgctg agacatatag gcggtcctga gagagaacct
                                                                      240
ccccaggcac ttcggccacg ggatagaaga cggcaggagg aaattgatta tagacctgat
                                                                      300
ggtggagcag gtgatgccga tttccattat aggggccaaa tgggccccac tgagcaaggc
                                                                      360
ccttatgcca aaacgtatga gggtagaaga gagattttga gagagagaga tgttgacttg
                                                                      420
agatttcaga ctggcaacaa gagccctgaa gtgctccggg catttgatgt accagacgca
                                                                      480
gaggcacgag agcatcccac ggtggtaccc agtcataaat cacctgtttt ggatacaaag
                                                                      540
cccaaggaga caggtggaat cctgggggaa ggcacaccga aagaaagcag tactgaaagc
                                                                      600
agccagtcgg ccaagcctgt ctctggccaa gacacatcag ggaatacaga aggttcaccc
                                                                      660
gcagcggaaa aggcccagct caagtctgaa gccgcaggca gcccagacca aggcagcaca
                                                                      720
tacageceeg caagaggtgt ggetggacca egtggacagg ateeggteag cageceetgt
                                                                      780
ggctagagga acaccagcac aaacgacagc ctcaagtctc cttcgagctt tgtaaagtaa
                                                                      840
tctcacaagc tgccaaagaa cacatcccaa gcagtaatca cagatgatga agcaccttac
                                                                      900
aggacccctc caccctcaaa cgcgcatgtc cagagaagta ttaatgcctt aatagactac
                                                                      960
tgaaggttaa ctatttacat cattctaaaa atattttaag ctgtattaca gtgctataaa
                                                                     1020
ttctcctttt aaggaaaaaa gtaatagatt gtttccttac cgtgtttgaa aggaaaaaaa
                                                                     1080
ataaatccat agtaactaaa aaaaaaaaaa aa
                                                                     1112
<210> 100
<211> 887
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (303)
<223> n equals a,t,g, or c
<400> 100
```

ggcacgagag aattataggt gcatgatggg gcttttggag actggcaatg ttctgttttg	
333 - 3399 - GCCC GCGCGCCC ACETTATOCT AACTCAFFA ACETTATOCT	60
John To good Coloco alalyly aleger caa age at at a sate and and a sate and a	120
stands conducted additional panagraphy casttagang complete	180
teresteta aagteetatt teetetette agtetateat taetaetagt teetagtata	240
a conscious addudacaar dardratata tototogo	300
ttctttttgt aattacaaaa atataaactc actacatatg ttattctacc acatgctttt	360
tttcatgtaa cagtatgtct tagacatctt ttcctattat tgcgtggaag ctatcaacct	420
tattetttgt aatgacteca taaattatte taetgtgaaa acacaccatg tttaaccagt	480
tctctgttag tgaacattta ggatttttcc agtttttaaa tattacagtg acatattaaa	540
cattaaacat atcttttgac acatgtgctt ggaggethat tattacagtg acatattaaa	600
cattaaacat atcttttgac acatgtcctt gcacacatgt ataggtatga tggactttaa	660
caccetting charactet tageacataa egiaaatate ceatagagte aaaaccacce	720
ttaaaacttc ctcaggaggc tgggtgtggt ggctcacgcc cgtaatccca gcactttagg	780
aggccgaggt gggcggatca cgaggtcagg agatcgagac catcctggcc aacatggtga	840
aaccctgtct ctactaaaaa tacaaaaaaa aaaaaaaaaa	887
<210> 101	•
<211> 1248	
<212> DNA	
<213> Homo sapiens	
<400> 101	
ccacgcgtcc ggtctattaa ggtgacagaa ctgaagggcc tggccaacca tgtggttgtg	60
The state of the s	120
The state of the s	180
The description of the contract of the contrac	240
The standard of the standard s	300
o o o o o o o o o o o o o o o o o o o	360
	420
	480
galuaduad dadaccaraa ccccaataat amaataa	540
The mongroup of the contract o	600
TO TO TO THE THE TOTAL TO THE COLOR OF THE TOTAL TOTAL TOTAL TO THE TOTAL TOTA	660
The state of the s	720
	780
The desired additional carage of the services	840
Jan	900
The strong court of the strong contract of th	960
The description of the second	1020
The state of the s	1020
	1140
Ctddddddddd AAAAAAAAA aaaaaaaaa a	1200
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	1248
\cdot	
<210> 102	
<211> 1841	
<212> DNA	
<213> Homo sapiens	
<400> 102	
ggcacgagct ggtgcaggag ctggagcagt accagttgtt gccgaagaga ttggactggg	60
	120
TOTAL TOTAL CONTROL CONTROL TOTAL AFFACTION	180
The state of the s	240
TO TO TO THE DUAL COURT OF THE PARTY TO THE TARK	300
gtggactata aaattetetg catttageat agaacattga attttattae ataagtgttt	360
The second decaying the se	420

```
aatgggtcct tctgcattaa aaaagttgat aaagatatag atgttaagct aaaaagcttg
                                                                        480
 aatttattct gagatgttgt ttgaggtttt atcaatgtga gttcttaatt ttccattttg
                                                                        540
 tageteacte ataaetteat tgaatgtaat tgtatgtgte taataaacte tttaaatget
                                                                        600
 agaggtattt catttgatat ttagcttctg aaagcccttt gttataaagg gaagtgattt
                                                                        660
 tctaaatttg gtagcagett gaaaacaage accacaaatg taaaattgca tggttettt
                                                                        720
 tttactaatt taattttttt gccaagttat gtgctagcaa aatgtttgaa attgctattc
                                                                        780
 tttttctgtt agcttggtat atatgactta gactaagtat atgtatgtga tataatatat
                                                                        840
 gtgtacatgt ggctaaactg ctaatgggca aagattgttg gttcttctga tgctgagtat
                                                                        900
 acatcaagtt catcccatga ttatttttta ttacgagcta ttgctgatgt gatgctactc
                                                                       960
 ttagctgcac caccagtata aagcttaact gagaaagaac acccctccct cttagttccg
                                                                      1020
 tgccagactg gtttgttagc tttagttgtt tcttagtgct ctagccatgt caccctgtgg
                                                                      1080
 cactaatcag gacactgttt atattettee tettetgtgt ecetatteae tttgtaatta
                                                                      1140
 ttctatttta gctggttata tggtagttaa ttggtagcct gttttccttt gttttcttta
                                                                      1200
 cttgtgtcaa gctccatgtc gttgttttgt gatgagcagt gatttagaac tggccagacc
                                                                      1260
 ttagaagagt cttcagcaag ctgccctagg gtaggggcat attctctgtt tgttttgtac
                                                                      1320
 tggtacccca ggaggatgcc ttgtaactgc tctttggtca tgatagtggt aatgaatcca
                                                                      1380
 ttcatgtaaa atgcaatatg tctgttagtt tggggcaccc gaggctcagt agaggtttag
                                                                      1440
 tatgatgcct tccctttctt ctttgagtta tgaagaaact aatttctggc acagtgtgac
                                                                      1500
 tgcttcaaac ctatgttctg ctgctatctc ctatagagga gggcagactt ttttttctta
                                                                      1560
gtgcatttgc tggtaggtca aggagggatg gactttgaat tagtgttcag agtcacagat
                                                                      1620
cctacgtatg tgctcttcag tagaggattt tctgtgatcc tacaatgaag ggaaagctat
                                                                      1680
ataatttatt ggtacttcct aaagatagag actaaagtca tggtagtatt ggccaaatta
                                                                      1740
attttaaaac tgaacaagtc atgaattttc atataaatca aagaactttt tccttataaa
                                                                      1800
agcttccatt tttgaaaagt tgaaaaaaaa aaaaaaaaa a
                                                                      1841
<210> 103
<211> 685
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (678)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (679)
<223> n equals a,t,g, or c
<400> 103
actacggctg cgagaagacg acagaagggg ggcggcgacg gaggaggagg atggaggcgg
                                                                       60
tggtgttcgt cttctctctc ctcgattgtt gcgcgctcat cttcctctcg gtctacttca
                                                                      120
taattacatt gtctgattta gaatgtgatt acattaatgc tagatcatgt tgctcaaaat
                                                                      180
taaacaagtg ggtaattcca gaattgattg gccataccat tgtcactgta ttactgctca
                                                                      240
tgtcattgca ctggttcatc ttccttctca acttacctgt tgccacttgg aatatatatc
                                                                      300
gatacattat ggtgccgagt ggtaacatgg gagtgtttga tccaacagaa atacacaatc
                                                                      360
gagggcagct gaagtcacac atgaaagaag ccatgatcaa gcttggtttc cacttgctct
                                                                      420
gcttcttcat gtatctttat agtatgatct tagctttgat aaatgactga agctggagaa
                                                                      480
geogtggttg aagteageet acaetacagt geacagttga ggageeagag acttettaaa
                                                                      540
tcatccttag aaccgtgacc atagcagtat atattttcct cttggaacaa aaaactattt
                                                                      600
660
aaaaaaaaa aaaaaaanna aaaaa
                                                                      685
```

<210> 104 <211> 1168

<212> DNA

```
<213> Homo sapiens
<400> 104
ggcacgagec eccecect eccteetee ecteettee ecggeecegg etetggeece
                                                                         60
ggcccattcg ctgttgggtc ttctgctagg gaggatgtcg ggttcgtcgc tgcccagcgc
                                                                         120
cetggeeete tegetgttge tggtetetgg eteceteete ceagggeeag gegeegetea
                                                                        180
gaacgtgaga gtacaatctg gacaggatca gaagtagaga atgaagttgt aagagaaagg
                                                                        240
gaaagacaga agaaaggctg cagtagtaca aggagaaaag caggatgcaa gaatgaggaa
                                                                        300
tgaatctttg tttgaggagc atccggaaaa atataagctg tcagaaagag taaatagacc
                                                                        360
agggacctct aaagtaaatt cacacatcaa agttaaaata atgttggaga atcacctcct
                                                                        420
gtgaaaaatt ggcttagctt tcagtatgcc tctttaaaca aaacattatc attttataca
                                                                        480
aatttttaaa atgttgttca taatatggag tttaatttta agcaccttag ataatgcatt
                                                                        540
tgttagcttg tttgtgaaat ttttatggat tttttttttc aaaattccta attttagttg
                                                                        600
gtaaggatta acttcgggaa gacaggaaac ccctccagta aaattaattg gttataaatg
                                                                        660
gttacatatt ttaggcttat atacataaca aaacttctaa ctgaatttaa aagtgatctt
                                                                        720
gtgtaaaaac attattctag tgatattgat gtctaaattt aaaagagtga atacacaagt
                                                                        780
aaaatatatc tgctttttaa ggatctattc agtttaggaa ggagagtcaa aatttgtatt
                                                                        840
taatttcaat ttattatatt gtctgaaact gaaagtaggc ctaacttttg tttgcttttg
                                                                        900
tgtatgtaca aaggcaaaca tttatcaatc aactcttatt aatttgagat tattttgacc
                                                                        960
tgattgctca gaacttttgt ctatctgtat agaaatgggg ttttctaaat atttaaagaa
                                                                       1020
tttcctacgt atgtaaattt acttgcacac agtaagttag gggagacatt taatattcct
                                                                       1080
cagatacctg ctgttttctt gttgtgtttc ttgtttttca aataaataaa ctgagtgtta
                                                                       1140
tctgttcatt aaaatagaat gtagtcag
                                                                       1168
<210> 105
<211> 1175
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (24)
<223> n equals a,t,g, or c
<400> 105
gcatggaaat atatcctata catnaagtgc tgtcctgagc tgtacatwta gccatattta
                                                                         60
grattttatt argcaggatt caaggcctcc taaaaatgga tattcagata tttaaaaata
                                                                        120
cactattaag gaatgcaatg taaaatcaca ctaacatttg catagcactt tatagtttac
                                                                        180
aaaatgcctt cacatatacg agctcattta tttctccttc tcttcttct attcatctat
                                                                        240
caaggaatat cttctataag ccaggcatca ggcttgacac tgaagacaca aaatgaaaaa
                                                                        300
gatatacaag tototattot taaagagtto gtagtttaat gaagtgacco aaaaagtaaa
                                                                        360
cagttgcaat ataacatgca aaggagtgtg atgaaataca cctagagtac tatgaaaatg
                                                                        420
gcatagaggg accaagcctg taatgaagtc agagaaagta aaagaggtga agcckgagct
                                                                        480
aagttttaaa ggatgaacaa aaattgacca gacaaaggag ttggagaagt gggatggtac
                                                                        540
aggataaaag cctttctaat ttggggaagt gtgagtcgtc gatattagat gatattatta
                                                                        600
atagattgtg cacagagtgc acatggtaga gaggtgagag atgaggtagc actgttaggg
                                                                        660
cagataggtg aattotcaca gtaaccotgt gwtgtacmtt atcatcotcc totccattat
                                                                        720
gtaggggaga acaccaaagc ccaaagagat taagtaactt tccagaattt tgatttcaaa
                                                                        780
tcctaagcgt ttggtagctg tgtttccctg catatagtct gcctcaaaag aaaatttaaa
                                                                        840
gttgtatttt gtattaaatt taaacaaata raggccgggt gcagtggctc acgcctgtag
                                                                        900
tcccagcact ttggtaggcc aaggtgggcg gatcacctga ggtcaggatt tcgagaccag
                                                                        960
cctggccaat atggtgaaac cccatctcta ctaacataca aaaattagct gggcatggtg
                                                                       1020
gcacatgeet gtaateetgg ctaettggga agetgaggea ggagaatete ttgaaceeag
                                                                       1080
gaggcggagg ttgcagtgag ccgagattgt gccactgcac tacaacctgg acaacagagc
                                                                       1140
aagattccgt ctcaaaaaaa aaaaaaaaaa ctcga
                                                                       1175
```

```
<211> 1021
  <212> DNA
  <213> Homo sapiens
 <400> 106
 ggcacgagaa taaaatagat ttagggtatt cgacacattt aatacataat gaatagtgaa
                                                                       60
 tttcttgaat aatataatgc ctaaatgtga ctgatatgca aaatatttaa atcccttaca
                                                                      120
 ggatgtgcat ttaccaatca gaacagatgt tggctttact gcttgtatta gttttctgta
                                                                      180
 ttagtttgct tgtattagtt tgctggggca gccataacaa agtaccacag aaatttatat
                                                                      240
 tetcacaatt etgggggeta gaagaetgag ataaagatgt etgeagggtt ggtttettet
                                                                      300
 aaggactete tigitggett gtagatggtt gtigieteee igigietica ceigeactie
                                                                      360
 cctctgtacc tgcccgtatc ttaatctctt cttataagga cagagtcata ttggattagg
                                                                      420
 gcctcaccct aatgacctcc ttttaactta attacctctt tagagactct gtcttcaaat
                                                                      480
 acaggcacat tetgaggtae tttaaggtta gaaetteaae atatgeattt ggaaetgggg
                                                                      540
 cataactgag tccatgacac tattggagag ggttctggag ccataacaga tggaatgaat
                                                                      600
 gggccttgtt tggattctaa tttaaataaa ccaactatag aaaaacatgt tttaggcaat
                                                                      660
 cagatttttt gttatgggct aaatatttag aagatgttaa gaaatgttta ttaatttcat
                                                                      720
 taggtaaaat aatgttggtg tggttatgta gagtatgact tacgtgcaca gaacagaatg
                                                                      780
 ataaggtgtc caggatttgc tttaaaatat tctagatttt tttaaaaggt gagaggaaat
                                                                      840
 gacacatgta tatctaaatt ttgatggttg ttgaagctgg ctggtaggta cctagtattt
                                                                      900
 cattatactt tttcctttgt gtgtttggaa atttttataa taagcaggaa aaataaaagt
                                                                      960
 1020
                                                                     1021
 <210> 107
 <211> 830
 <212> DNA
 <213> Homo sapiens
<400> 107
ggcacgagta aggactgtgt tetttatgca tttettgate caggeatgge agtteetett
                                                                      60
ttcctgtaca tattcacact cctgccactt ctaccctttc tcttatccct ctgcttttca
                                                                     120
cctctgactg taaaaagaag tagcagttcc gaaagcaaga gttccctatg aacacggaag
                                                                     180
aagacattgg caacttttga gtacaacaac tatatttaat agagtaattt aagaacatca
                                                                     240
gccagtgaat tttatacaag atagtgaaag agaaaaggaa gattaattag gggtagttta
                                                                     300
ggatgccatt aaatagccta gaattagggg agtagtcgtt gaatagaaag gaggccacaa
                                                                     360
atttgaggga tataagctaa gaattggtaa gccaagaaga aggaaaaggt ttgggcagta
                                                                     420
aggataatga ggaacaaaat agagaactca gaagcaatat ctgactgtta tcattggaag
                                                                     480
aatttttttg cttgcttgag gctggatatt gaagtggatc aggatacttg agtgactatc
                                                                     540
tgatgggctt ttggaactag ctctcaagag gtgaaaatta gctttttttt ctttttcttt
                                                                     600
cttttttttt ttttttgagg caaggtctca ctgttgttga ggctgaacct cctgggctca
                                                                     660
agcagttgtc ccattgcagc ctcctcagat actctgtaag ccaaggcagg gggaatattt
                                                                     720
tgtgctcagt agtttgaggc tgtggtgagc taagatcaca ctgctgtgct cacttcagcc
                                                                     780
830
<210> 108
<211> 1301
<212> DNA
<213> Homo sapiens
<400> 108
aagagggtgg ctgccgagtt cttgtttgtc ctgtgctctg agagtggtga gttatggaga
                                                                     60
cccaggtccc agcccaccc acctcagcca ggcttgcaca ctgacctctg ccttgcccct
                                                                    120
cagtgccccg attcatcaag tacacaggct atgggaatgc tgctggcctt ctggctgcca
                                                                    180
ggggcctcat ggcaggaggc cggcccgagg gccagtactc agaggatgag gacacagaca
                                                                    240
cagatgagta caaggaagcc aaagccagca taaaccctgt gaccgggagg gtggaggaga
                                                                    300
agccgcctaa ccctatggag ggcatgacag aggagcagaa ggagcacgag gccatgaagc
                                                                    360
```

```
tggtgaccat gtttgacaag ctctccagga acagagtcat ccagccaatg gggatgagtc
                                                                         420
 cccggggtca tcttacgtcc ctgcaggatg ccatgtgcga gactatggag cagcagctct
                                                                         480
 cctcggaccc tgactcggac cctgactgag gatggcagct cttctgctcc cccatcagga
                                                                         540
 ctggtgctgc ttccagagac ttccttgggg ttgcaacctg gggaagccac atcccactgg
                                                                         600
 atccacaccc geoeceactt etecatetta gaaacceett etettgacte eegttetgtt
                                                                         660
 catgatttgc ctctggtcca gtttctcatc tctggactgc aacggtcttc ttgtgctaga
                                                                         720
 actcaggctc agcctcgaat tccacagacg aagtactttc ttttgtctgc gccaagagga
                                                                        780
atgtgttcag aagctgctgc ctgagggcag ggcctacctg ggcacacaga agagcatatg
                                                                        840
ggagggcagg ggtttgggtg tgggtgcaca caaagcaagc accatctggg attggcacac
                                                                        900
tggcagagcc agtgtgttgg ggtatgtgct gcacttccca gggagaaaac ctgtcagaac
                                                                        960
tttccatacg agtatatcag aacacacct tccaaggtat gtatgctctg ttgttcctgt
                                                                       1020
cctgtcttca ctgagcgcag ggctggaggc ctcttagaca ttctccttgg tcctcgttca
                                                                       1080
getgeceact gtagtateea cagtgeeega gttetegetg gttttggeaa ttaaacetee
                                                                       1140
ttcctactgg tttagactac acttacaaca aggaaaatgc ccctcgtgtg accatagatt
                                                                       1200
gagatttata ccacatacca cacatagcca cagaaacatc atcttgaaat aaagaagagt
                                                                       1260
tttggacaaa aaaaaaaaaa aaaaaaactcg a
                                                                       1301
<210> 109
<211> 1932
<212> DNA
<213> Homo sapiens
<400> 109
 aattttttt tttttttt ttttttgaa aaaaaaaatg ggtagtgtat attttgcagg
                                                                          60
 tttaagacaa ctcaggacaa taaaaacaat ggactttaca tgtgtatata tatagctctc
                                                                         120
 ttaggcacca taatcagtat gagccaacaa tatttaaact tgattcaggc cacattcaga
                                                                         180
 catttgctct tatatacaaa tatttaaatt aaatacaatc tgaaatgtgt tctgttacat
                                                                         240
 acaaaaaagg aaaaactata caacgcagag cagtgtgtgt gttttaaata attacattta
                                                                        300
 catgtaagct aaatggaacc agcaatggtg ctcaagtttt tatcatccct tccagaaaat
                                                                        360
 ctttttctac catctcttct attttttgcc tggctttgct ggaacatggt ttgtggttct
                                                                        420
ccagtttcat gtccttatta gggaaggcat ttgagtagag gataggactc cctgagtgtc
                                                                        480
ctccacatcg gcttgtgact ttgctgttga agacttgact gagcacattg aagaacggca
                                                                        540
 ggagetgete catactgege aeggtgeaga tggtgageag caagtgeeet ggeteecaae
                                                                        600
ccaatgttct ccctgagttg tettectetg gatttttetg cagaaaacaa aaagtgaact
                                                                        660
ggtattaata caacagacaa tgtggtatgt tagaaaaatt aaaaatatat aaactttggc
                                                                        720
aattggtcaa gaaatgaata caaatgacat taagtttcta actcctgacc tgatcaaaac
                                                                        780
ccttggtgct tctgagacct tttactgcca tttattagtt ttacatggag cagtctaaca
                                                                        840
ttgtagtaat agttcccaac tagaatgcgc agataagctt agttaacaga aatagctttg
                                                                        900
aacaggaata gagtcaaaca taaaagtttt atgttgtgct ttgtatttac tcaaaaagct
                                                                        960
cccaggtttc tgaaccctca ctactgtaac caaggactag gtcacaaaat tactacagaa
                                                                       1020
aaaaggaaca aagtgcttta tacatttcat aatatatccc cttttattat aattagttaa
                                                                       1080
ttccctttta tctaaatggc ctaaatttgc catgatggta gcagtgtcca aagtgaataa
                                                                       1140
ttactgtcag tactgcatca cagagaaagg aagggateee teaggagaca etgetgtete
                                                                       1200
cttctgggtt gtgctaaaca acatagggag gaaagctgga cctggagtca aaggaattga
                                                                       1260
gttagtgtgc tggctctgcc atacttacgg caccettggg caggatatac aaaggttcct
                                                                       1320
cacttataaa atgggacagt ctaaaactac cttttagtag agaagtcaaa tgagaaggta
                                                                       1380
tgtgaaaact ctgtcaacta aatataaaga ctaataattt gggtattaag aggctagttt
                                                                       1440
gagaagccac ctgaattaca caaacacagc tacagacatc attctgtcta gagaaagata
                                                                       1500
agagagaaca ggttggttga acttgggcag aatcacagat acaattccac actaaagaat
                                                                       1560
gaaaataagc aatgaactag acagaaggaa gaaatcatga agacttagga agcagaatta
                                                                       1620
caatctgtca tattaacaaa tggagtttgc cttctaagat cagatgttgc tcagaaactt
                                                                       1680
tcattgttta cctaataatt taatatcact agtttcctag tgggtcaagc agatgcaaaa
                                                                      1740
tocagottat tttottotat gtgototcaa gottattgot tattttaaag taaaatootg
                                                                      1800
aaaaaggaaa atattaggtt ggtgcaaacg taattgcggt ttttgcattg ttgaaatttg
                                                                      1860
ccgttttata ttggagtaca ttcttaaata aatgtggtta tgttatacaa aaaaaaaaa
                                                                      1920
aaaaaactcg ag
                                                                      1932
```

<223> n equals a,t,g, or c

```
<210> 110
  <211> 1534
  <212> DNA
  <213> Homo sapiens
 <220>
 <221> SITE
 <222> (1212)
 <223> n equals a,t,g, or c
 <400> 110
  togaagatag gttcgagacg gtggatgttg cagctgatca tgcagttggg ttcggtgctg
                                                                            60
  ctcacacget geeeettttg gggetgette ageeagetea tgetgtaege tgagaggget
                                                                           120
  gaggcacgcc ggaagcccga catcccagtg ccttacctgt atttcgacat gggggcagcc
                                                                           180
  gtgctgtgcg ctagtttcat gtcctttggc gtgaagcggc gctggttcgc gctgggggcc
                                                                           240
  gcactccaat tggccattag cacctacgcc gcctacatcg ggggctacgt ccactacggg
                                                                          300
  gactggctga aggtccgtat gtactcgcgc acagttgcca tcatcggcgg ctttcttgtg
                                                                          360
  ttggccagcg gtgctgggga gctgtaccgc cggaaacctc gcagccgctc cctgcagtcc
                                                                          420
  accggccagg tgttcctggg tatctacctc atctgtgtgg cctactcact gcagcacagc
                                                                          480
  aaggaggacc ggctggcgta tctgaaccat ctcccaggag gggagctgat gatccagctg
                                                                          540
  ttettegtge tgtatggeat cetggeeect ggeetttetg teaggetact acgtgaccet
                                                                          600
  cgctgcccag atcctggctg tactgctgcc ccctgtcatg ctgctcattg atggcaatgt
                                                                          660
  tgcttactgg cacaacacgc ggcgtgttga gttctggaac cagatgaagc tccttggaga
                                                                          720
  gagtgtgggc atcttcggaa ctgctgtcat cctggccact gatggctgag ttttatggca
                                                                          780
  agaggetgag atgggeacag ggagecaetg agggteacce tgeetteete ettgetggee
                                                                          840
  cagctgctgt ttatttatgc tttttggtct gtttgtttga tcttttgctt ttttaaaatt
                                                                          900
 gttttttgca gttaagaggc agctcatttg tccaaatttc tgggcttcag cgcttgggag
                                                                          960
 ggcaggaacc ctggcactaa tgctgtacaa ggttttttc ctgttaggaa gaacttgagg
                                                                         1020
 ccagctgccc actgagtctt ctgtccctga agaaagggag tattgggcag ggcttgggat
                                                                         1080
 ccggctactg agagtgggag agtgggagac agaggaagga agatggagat tggaagtgag
                                                                         1140
 caaatgtgaa aaattcctct ttgaacctgg cagatgcagc taaactctgc agtagtgttt
                                                                         1200
 ggagactgtg anagggagtg tgtgtgttga cacatgtgga tcaggcccag gaagggcaca
                                                                         1260
 ggggctgagc actacagaag tcacatgggt tctcagggta tgccaggggc agaaacagta
                                                                         1320
 ccggctctct gtcactcacc ttgagagtag agcagaccct gttctgctct gggctgtgaa
                                                                         1380
 ggggtggagc aggcagtggc cagctttgcc cttcctgctg tctctgtttc tagctccatg
                                                                         1440
 gttggcctgg tgggggtgga gttccctccc aaacaccaga ccacacagtc ctccaaaaat
                                                                         1500
 aaacatttta tatagacaaa aaaaaaaaaa aaaa
                                                                         1534
<210> 111
<211> 2871
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (1234)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1259)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1283)
```

```
<220>
<221> SITE
<222> (1284)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1287)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1378)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1912)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1913)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1935)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (1947)
<223> n equals a,t,g, or c
<400> 111
gcggcgaggc ggcggtggtg gctgagtccg tggtggcaga ggcgaaggcg acagctctag
                                                                          60
gggttggcac cggccccgag aggaggatgc gggtccggat agggctgacg ctgctgctgt
                                                                         120
gtgcggtgct gctgagcttg gcctcggcgt cctcggatga agaaggcagc caggatgaat
                                                                         180
ccttaggatt ccaagactac tttgacatca gatgagtcag taaaggacca tactactgca
                                                                         240
ggcagagtag ttgctggtca aatatttctt gattcagaag aatctgaatt agaatcctct
                                                                         300
attcaagaag aggaagacag cetcaagage caagaggggg aaagtgtcac agaagatatc
                                                                         360
agctttctag agtctccaaa tccagaaaac aaggactatg aagagccaaa gaaagtacgg
                                                                         420
aaaccaggta gtctggacat tttccttgct ttttgattta tttaggggac aactgaaaat
                                                                         480
tttaagctaa tgaataaaga agctgaagaa gactggcttc actgattatt acccacaaat
                                                                         540
aatatatgga gtgtagtttg gagagaattt ctagatttta atataccaaa gttattcact
                                                                         600
taacagattt gacccaagtt acataagctg aattccatag atgagattgc atttatcttt
                                                                         660
ttattaatat cctgacgttt ctttttggaa tcaagtatgt aaataccttt ctcactgaga
                                                                         720
aatttaaaaa aataatatgt cagggtagca tacaaaacag aaactcagga agaactgtct
                                                                         780
aacatgccac tatggttatt ttcaggattg tggtaggatt attactacaa taaccttgtc
                                                                         840
gatcatecta ggtaatttga tttaggtgtt tateeeteea gettgaeget tgtgtgtaet
                                                                         900
tocatoctat totaactttc acaatottga aacototact titcaattta aaacttgaco
                                                                         960
tttttcacaa tagtottota aatattaato ottttaaaaa ttattaatgt atgoattota
                                                                        1020
ccatgtcttg ttgcccttct gtcatttatt atttcaatga tcactgcctt atctgtttat
                                                                        1080
tattccatta ctaaccactc tcaggaactg atccaaattg gaacttttta aaaaaataga
                                                                        1140
attttttttt actacaata taccattctc atgaatagaa aactatttt aaaaaattaa
                                                                        1200
aagtacctga atcccatcaa ctagaaataa cagntgctca catcatggag aatattctnt
                                                                        1260
caggattttt ttaggtgtat cannagnttt aagaaaagta aattgggtca aaatgtgtat
                                                                        1320
tctatgttgc agcctgcatt ttgcacttaa cagtttacca tgaatgtttt tccatgtnat
                                                                        1380
```

```
ttagcttatt ttcaatttga tagttaatgg ctataaaaaa ttttatttgt aaatatatgg
                                                                        1440
  taccataaac caaaacgitt atgittigci gggagatcat titagatgia ictitigigcc
                                                                        1500
  tatctgtgaa aatttcctta gaattcttaa agtaaatttg ctgaattgga tgtaaaagtt
                                                                        1560
  taaaagacag tagatacaca ttgttaaatt gttcttcaga aaagtgattc tgtttacttc
                                                                        1620
 ctctgaaaat acctctttcc ctgtacctca cccaagtgtg tattaccatt ttaaaatttt
                                                                        1680
  tactaaatat aatttaattc attottoatt tittaattag ggatititti tittittitt
                                                                        1740
  tagtttttat atttttagag acatggtctc actctgtcac tcaggctgga gtgcagtggc
                                                                        1800
 acgatcatag ttcactgcag ccttgaactc aggtgatctc cctctttggc ctcccaaagt
                                                                        1860
 gctgggatta caggcatgag acactacatc aagcctggga aatttttaaa cnngcataaa
                                                                        1920
 aagtagaggg taaantgcca ccccttncca accctcactc agcttcaata aacatcaaca
                                                                        1980
 ttctgctggt cttttcatct ttactaccct acacacattt gccctttttt tctttcctag
                                                                        2040
 aatattttaa ggcaaattcc gcatgtgctt ttacatctct gtgtctctct ctagtaagaa
                                                                        2100
 tgctcttttc agcaaaatca cagaattagc cattactcct ttattaccta atacccagac
                                                                       2160
 tgtttaattt ttctagttgc ttcaagtgtc tgtttacaat ttatttgttt gaattcattt
                                                                       2220
 ctaagctact caggaggctg agtcaggaga attgcttgaa ccctagaggc agaggttgca
                                                                       2280
 gtgagctgag atcgtgccac tgtactccag cctgggtgac aaagtgagac tccatgtcat
                                                                       2340
 aacaaaacaa aacaaaacaa aacatgacga aacaaacaac acaattgaat aatatccctc
                                                                       2400
 agttattatt tagagaagag atgactcaat ataaatgtaa tatgtgatcc atttggtcac
                                                                       2460
 agaatgaatt tcagtcattg cttaggtaag cctggtaaaa aaaaaaaaa cgggattcca
                                                                       2520
 tgattagtgt gttgaaaaag gggtttgccc aaagtgtaac agacagagcc gggtgcagtg
                                                                       2580
 getcacgeet gtaateecag aactttggga ggecaaggea ggeggateae gaggtcaaga
                                                                       2640
 gattgagaca atcctggcca acatggtgaa gccctgtatc aactaaaaat acaaaaacaa
                                                                       2700
 ttagctgggc gtggtgacat gtgcctgtag tcccagctac tcaggaggct gaggcaggag
                                                                       2760
 aattgcttga acctgggagg cggaggctgc agtgagctga gattgctcca ctgcactcca
                                                                       2820
 2871
<210> 112
<211> 1037
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (936)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (946)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (951)
<223> n equals a,t,g, or c
<400> 1:12
ccggtccgga attcccgggt cgacccacgc gtccgggcgg gcgcaggacg tgcactatgg
                                                                        60
ctcggggctc gctgcgccgg ttgctgcggc tcctcgtgct ggggctctgg ctggcgttgc
                                                                       120
tgcgctccgt ggccggggag caagcgccag gcaccgcccc ctgctcccgc ggcagctcct
                                                                       180
ggagcgcgga cctggacaag tgcatggact gcagcacctc ctgccccctt ccggctgctt
                                                                       240
tggcccatcc ttgggggcgc tctgagcctg accttcgtgc tggggctgct ttctggcttt
                                                                       300
ttggtctgga gacgatgccg caggagagag aagttcacca cccccataga ggagaccggc
                                                                       360
ggagagggct gcccagctgt ggcgctgatc cagtgacaat gtgccccctg ccagccgggg
                                                                       420
etegeceact cateatteat teatecatte tagagecagt etetgeetee cagaegegge
                                                                       480
gggagccaag ctcctccaac cacaaggggg gtgggggggg gtgaatcacc tctgaggcct
                                                                       540
gggcccaggg ttcaggggaa ccttccaagg tgtctggttg ccctgcctct ggctccagaa
                                                                       600
cagaaaggga gcctcacgct ggctcacaca aaacagctga cactgactaa ggaactgcag
                                                                       660
```

PCT/US99/17130

75

```
catttgcaca ggggaagggg gtgccctcct tcctagaggc cctgggggcc aggctgactt
                                                                      720
 ggggggcaga cttgacacta ggccccactc actcagatgt cctgaaattc caccacgggg
                                                                      780
                                                                      840
 gtcaccctgg ggggttaggg acctattttt aacactaggg ggctggccca ctaggagggc
 tggccctaag atacagaccc ccccaactcc ccaaagcggg ggggggatat ttattttggg
                                                                      900
gagagtttgg aggggagggg gattttttt ttaaangatt ttttantttt naaaaaaaa
                                                                      960
 1020
aaaaaaaaa aaaaaaa
                                                                     1037
<210> 113
<211> 2214
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (289)
<223> n equals a,t,g, or c
<400> 113
cggacgcgtg ggtcgaccca cgcgtccggg aaaaarggaa aaratgccgt gtaaaatctc
                                                                      60
gttctgtgtc tgaattgccg taggctcaga tcttcatttg aggttctgtg tctgaattgc
                                                                     120
Cgtaggctca gatcttcatt tgaggttatg ttctataagt taacgttgat cttgtgtgag
                                                                     180
ctttcggtag ctggagtaac acaggcgcc tcacagcgac ctctccagcg ccttccaagg
                                                                     240
cacatetgca gecagegtam tecteetggg agatgeetee teaaggeent getecagaee
                                                                     300
acgtgggrar ggcctgacaa gccaattccc aggctgtccc cacccttgra gagtgaccct
                                                                     360
aaacgctaga cagatgggga atgggaaaga aaagaaagct gcagacctca agttaaaatt
                                                                     420
CCCtcaaaaa cgtttttatt tatctgcttt ttctgaaagg ataaaggctt tttgaaaatt
                                                                     480
attitictaac aaataacatg aacacticta gaaaccctag aaaaacacaa agtaticaaa
                                                                     540
atagaaagaa aaattaccca ttactcttta agccagcatt atccattgcg gtgcttttgg
                                                                     600
agttgggtga ggccgtagcc tctgccaagt caaggagccc ggtggtggct gtggcattcc
                                                                     660
tgcagggttg ttttttttc tttgagatgg agtctcactc ttgtcacccc agctggaatg
                                                                     720
tggtggtgta aacagctcac tgcagccttg accctgaggc tcaagcgatc cttctgcctt
                                                                     780
ggcctcctga gtagctggga tcccaggcga gagtcaccac accctgtcca tgttcctgca
                                                                     840
ggtcttgata tgcgaggacg ctgtgtcttc cctgccacat tttcttcttc tttcttgaga
                                                                     900
cagaccettg ctccatcacc caggccagag tgtggtsgtg cgaacacggc tcactgcage
                                                                     960
ctcgaccctc aggctcaagc gatcctcacg cctcggaccc ccaaagtgct gggatcacag
                                                                    1020
gcgagagtca ccatgctggc ctgaatcttc agggtatttr cggttgargt gycacttact
                                                                    1080
tarccatscc tgtttcaaga gtgtaggtgg tcaccctgtc tctgccgctg acctggcctg
                                                                    1140
gaccctcggc tgtgagaggg aggggtgggc tgggctggag gaacctraag ccctcgtgat
                                                                    1200
gtcacaagcc catctggctg ggcatcccct gctgtgtcct gagctgcaca tgccccaggt
                                                                    1260
ggcccccaca gcagaggcga gccactgrag ggtgragggc ttccacggac ggtcttcagg
                                                                    1320
ggragaagaa gggcccaggc ccccaggaga ctcaggagac cagagcctgg ggtcaggggc
                                                                    1380
tmagcagggg ctyarccagg gctggatgtc cggagccagc cccgmagccc tgkgktcttt
                                                                    1440
gttcttcgca ctcccaccgt ccgtgtgaac agctccagcc ccacctgcgc ctccctgtgc
                                                                    1500
tgggctccat cagggagccc agaagacgtg tgtgcttctg aaattgggtc cctacatgcc
                                                                    1560
tttgtcccag tgcaccttgc tccttccatt tactatcgag atttaaatgc ctgttttctc
                                                                    1620
cccagaggtt gacggatata ttcagacgtt acgacacgga tcaggacggc tggattcagg
                                                                    1680
tgtcgtacga acagtacctg tccatggtct tcagtatcgt atgaccctgg cctctcgtga
                                                                    1740
agagcagcac aacatggaaa gagccaaaat gtcacagttc ctatctgtga gggaatggag
                                                                    1800
cacaggtgca gttagatgct gttcttcctt tagattttgt cacgtgggga cccagctgta
                                                                    1860
catatgtgga taagctgatt aatggttttg caactgtaat agtagctgta tcgttctaat
                                                                    1920
gcagacattg gatttggtga ctgtctcatt gtgccatgag gtaaatgtaa tgtttcaggc
                                                                    1980
attctgcttg caaaaaaatc tatcatgtgc ttttctagat gtctctggtt ctatagtgca
                                                                    2040
aatgctttta ttagccaata ggaattttaa aataacatgg aacttacaca aaaggctttt
                                                                    2100
catgtgcctt actttttaa aaaggagttt attgtattca ttggaatatg tgacgtaagc
                                                                    2160
2214
```

```
<210> 114
<211> 3300
<212> DNA
<213> Homo sapiens
```

<400> 114

/400> 114						
tcgacccac	g cgtccggtg:	a gaggcaagg	a cttttcatc	t ttagggcati	t ctgaccacgt	60
cergerrea	g agagattgti	cccggcgtc	t cagtgctate	g gggagcaggi	tectectaat	120
congetete	a ggtctcact	; tcttactgg	c tctgccagg	a tcagaagcca	a agaattctgg	180
agetteetgi	cctccatgc	: ctaaatatg	cagctgcca	c aacagcacco	actgtacttg	240
cyaagatgg	: tttcgggcca	ggtctggcag	g gacatactt	t catgattcct	ctgagaagtg	300
Lydagatati	: aatgaatgtg	, aaaccgggct	ggcaaagtg	c aagtataaac	catattotao	360
yaacaaagtt	ggaggttaca	i tctgtagctg	y tttggtaaaa	a tatactttat	tcaactttct	420
ggerggratt	atagattato	, atcatccgga	i ttgttatgag	g aacaatagto	aagggacgac	480
acagtcaaac	: gtggatattt	gggtgagtgg	ggtgaagcct	t ggatttggga	aacagctggt	540
acytataact	atgccatttt	cctacccaaa	cattaacato	; tcttcctgtg	atttttaggg	600
cagggtagtt	ctatccaggg	r gtaattttgt	cctctgtccc	aaggtcatct	gtcaatgact	. 660
ggggacactt	ttggttgtca	taccttgggg	gtgatgtgtc	, tgactggcat	ctggtggatg	720
gagaccaggg	n atacagetea	acatcctaca	gtgcccagga	cagcctccca	caatcaagaa	780
gracerages	ccatatgtco	atagagatag	agaaatacaa	atatagggg	aaagtgcctc	840
agctggcatt	caaagaccta	catcaagcac	ctagattctc	aatgccacac	gcaccttgta	900
gcaccttaat	aaatatcgtt	gccttctggg	ctccccactc	caacacttgt	gcatattccc	960
tatttcttac	atttcagtaa	gagtcaatct	aattagctta	attttttgg	aaggaaaatc	1020
tgagaagaaa	tggaagcaga	gaggactttq	caagaaggg	tactcaacta	attcaaagcg	1080
tggagttgag	catctggaat	gcgagttttg	cttctccagg	aaagggtcaa	atttctgaat	1140
ttgatatagt	ctatgaaacc	aagaggtgca	atgagacaag	ggagaatgct	tttctggaag	1200
ctggaaataa	caccatggat	atcaactgtg	ctgatgcttt	aaaaggaaac	ctaagagaga	1260
gcactgcagt	tgccctatca	cttatcaatc	tcttggggat	attctgaatg	catectitit	1320
tagtaaacga	aaagggatgc	aggaagtaaa	actgaactct	tacgttgtga	gcggcaccat	1380
cggtttgaag	gaaaaaattt	ccctctctga	acctgtgttc	ctgacttttc	gccataatca	1440
gcctggtgac	aagagaacaa	aacatatctq	tatctactaa	gagggatcag	aggaaggcca	1500
ctggtccacg	gagggctgct	ctcatgtgca	cagcaacggt	tcttacacca	aatgcaagtg	1560
cttccatctg	tccagctttg	ccgtcctcgt	gactettace	cccaaggagg	accetatact	1620
gaccgtgatc	acccaggtgg	ggctgaccat	ctctctacta	tgcctcttcc	tagccatect	1680
caccttcctc	ctgtgccggc	ccatccagaa	caccagcacc	tccctccatc	tagagetete	1740
cctctgcctc	ttcctggccc	acctcctgtt	cctgacgggc	atcaacagaa	ctgagcctga	1800
ggtgctgtgc	tccatcattg	cagggctgct	gcacttcctc	tacctggctt	acticaceta	1860
gatgctcctg	gaagggctgc	acctcttcct	caccatcaga	aacctcaagg	taaccaacta	1920
caccagcacg	ggcagattca	agaagaggtt	catgtaccct	gtaggctacg	agateceage	1980
tgtgattatt	gctgtgtcag	caatagttgg	accccagaat	tatggaacat	ttactcactc	2040
ttggctcaag	cttgataaag	gattcatctg	gagetteatg	gggccagtag	cartcattat	
cttgataaac	ctggtgttct	acttccaagt	tctatagatt	ttgagaagca	aactttooto	2100
cctcaataaa	gaagtttcca	ccattcagga	caccagagate	atgacattta	aaccccccc	2160
tcagctattt	atcctgggct	attettagga	ccttaatttt	tttatggttg	aagccattte	2220
gaagacgatt	ggatcaatca	ttgcatactc	attracrate	atcaacaccc	ttcaggagtagg	2280
gttgctcttt	gtggtacact	gtctccttaa	tcaccaaatt	cgaatggaat	ataaaaaata	2340
gtttagtggg	atgcggaaag	gggtagaaac	tgaaagcact	gagatgtctc	actataataa	2400
ccaaaccaaa	acqqaaqaaq	tagagaaatc	ctcagaaatc	tttcataaag	getetattat	2460
atcatcatct	qcagagtcaa	CCAAGCAACC	aceaccecea	gttcatctcg	tatatatta	2520
ttggctaaag	atgaactgac	ctaageaace	ccateggaat	gacccggaag	ttaggetge	2580
tttccgtttg	tctacagege	ccctataata	acacatacat	tggacaaatg	CLACCCCCCC	2640
tagettteet	gtgaaaagtc	tagggtgatt	acacacayat	ggctttttat	CCACLATTE	2700
agaacaagac	atttoogaga	attettagat	cacctatttt	gtagtgtggc	gttcatagaa	2760
aaqtqtcqqa	accegggaga	ttaaagat	ccagagteca	gragragrage	acgtgcaatg	2820
gctactgcca	Cottaccaca	aactcactca	cyggcgggag	aagtggattt	tcttcttgca	2880
ttctggcctc	tctactatat	tttstactad	clygcatctg	gattcagctc	atagttccct	2940
Cataattcaa	caattttcat	truatycecc	cadagatett	acattaacac	tccacattca	3000
aaaatgcatt	atcaggtagt	acygateagt	accaaagagg	gtgttgcatt	ttgcaataca	3060
gtagaactat	aaactacttc	yyayaygatg	Lygagaaata	ggaacacttt	tacactgttg	3120
5 - 555 ACCGC	addiagtic .	aaccatcgtg	yaagtcagtg	tggcgattcc	ccagggatct	3180

```
3240
 agaactagaa ataccatttg acacagctat cccattactg ggtatatacc ccaagggcta
 taaatcatgc tgctataaag acacatgcac acgtaaaaaa aaaaaaaaag ggcggccgct
                                                                        3300
<210> 115
<211> 1286
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (1149)
<223> n equals a,t,g, or c
<400> 115
 gactgcttta cggacattgg atgaagccga agcatttaga atggtgcctg gcacacagtt
                                                                          60
                                                                         120
 ggtgcgtgat atggttaagc tttgtgtccc cacccacatc tcatcttgaa tgtgacggtt
 tecceggete ectectgeeg ceatgtgaag aaggtegttg etteccette acettecace
                                                                         180
 accatgattg ccatggatgc tccccactcc aaagcagccc tggacagcat taacgagctg
                                                                         240
 cccgagaaca tcctgctgga gctgttcacg cacgtgcccg cccgccagct gctgctgaac
                                                                         300
                                                                         360
 tgccgcctgg tctgcagcct ctggcgggga cctcatcgac ctcatgaccc tctggaaacg
                                                                         420
 caatgeetge gagagggett catcaccaag gactgggace agecegtgge egactggaaa
                                                                         480
 atcttctatt tcctacggag cctgcatagg aactcctgcg caacccgtgt gctgaagagg
                                                                         540
 atatgtttgc atggcaaatt gatttcaatg gtggggaccg ctggaaggtg gagagcctcc
                                                                         600
ctggagccca cgggacagat tttcctgacc ccaaagtcaa gaagtatttt gtcacatcct
 acgaaatgtg cctcaagtcc cagctggtgg accttgtagc cgagggctac tgggaggagc
                                                                         660
 tactagacac attccggccg gacatcgtgg ttaaggactg gtttgctgcc agagccgact
                                                                         720
                                                                         780
 gtggctgcac ctaccaactc aaagtgcagc tggcctcggc tgactacttc gtgttggcct
                                                                         840
ccttcgagcc cccacctgtg accatccaac agtggaacaa tgccacatgg acagaggtct
cctacacctt ctcagactac ccccggggtg tccgctacat cctcttccag catgggggca
                                                                         900
                                                                         960
 gggacaccca gtactgggca ggctggtatg ggccccgagt caccaacagc agcattgtcg
 tcagccccaa gatgaccagg aaccaggcct cctccgaggc tcagcctggg cagaagcatg
                                                                        1020
 gacaggagga ggctgcccaa tcgccctacc gagctgttgt ccagattttc tgacagctgt
                                                                        1080
                                                                        1140
 ccatcctgtg tctqqqtcaq ccagaggttc ctccaggcag gagctgagca tggggtgggc
                                                                        1200
 aagtgaggnc cctgtaccaa gcgacttctg ccccggttca accctaccaa gcttggggga
                                                                        1260
acttactgca catagetetg acgttttgtt gtaataaatg ttttcaggce gggcaaaaaa
                                                                        1286
 aaaaaaaaa aaaaaaaaa aaaaaa
<210> 116
<211> 2189
<212> DNA
<213> Homo sapiens
<400> 116
ggaattegge acgaggegee egaggatgtg etgetggeeg etgeteetge tgtggggget
                                                                          60
                                                                         120
 gctccccggg acggcggcgg ggggctcggg ccgaacctat ccgcaccgga ccctcctgga
ctcggagggc aagtactggc tgggctggag ccagcggggc agccagatcg ccttccgcct
                                                                         180
 ccaggtgcgc actgcaggct acgtgggctt cggcttctcg cccaccgggg ccatggcgtc
                                                                         240
 cgccgacatc gtcgtgggcg gggtggccca cgggcggccc tacctccagg attattttac
                                                                         300
                                                                         360
 aaatgcaaat agagagttga aaaaagatgc tcagcaagat taccatctag aatatgccat
                                                                         420
 ggaaaatagc acacacaa taattgaatt taccagagag ctgcatacat gtgacataaa
 tgacaagagt ataacggata gcactgtgag agtgatctgg gcctaccacc atgaagatgc
                                                                         480
 aggagaaget ggteccaagt accatgaete caatagggge accaagagtt tgeggttatt
                                                                         540
 gaatcctgag aaaactagtg tgctatctac agccttacca tactttgatc tggtaaatca
                                                                         600
 ggacgtcccc atcccaaaca aagatacaac atattggtgc caaatgttta agattcctgt
                                                                         660
 gttccaagaa aagcatcatg taataaaggt tgagccagtg atacagagag gccatgagag
                                                                         720
 tctggtgcac cacatcctgc tctatcagtg cagcaacaac tttaacgaca gcgttcctgg
                                                                         780
 aatccgggca cgaattgcta tcaccccaac atgcccgatg cattcctcac ctgtgaaact
                                                                         840
```

1560 1620

1680

	gtgattttg	cctgggctat	tggtggagag	ggcttttctt	atccacctca	tgttggatta	900
				gtgctcctag			960
				ggactgaggt			1020
	aggaaatatg	atgctggggt	gattgaggct	ggcctctggg	tgagcctctt	ccataccatc	1080
	cctccaggga	tgcctgagtt	ccagtctgag	ggtcactgca	ctttggagtg	cctggaagag	1140
	gctctggaag	ccgaaaagcc	aagtggaatt	catgtgtttg	ctgttcttct	ccatgctcac	1200
	ctggctggca	gaggcatcag	gctgcgtcat	tttcgaaaag	ggaaggaaat	gaaattactt	1260
	gcctatgatg	atgattttga	cttcaatttc	caggagtttc	agtatctaaa	ggaagaacaa	1320
	acaatcttac	caggagataa	cctaattact	gagtgtcgct	acaacacgaa	agatagagct	1380
	gagatgactt	ggggaggact	aagcaccagg	agtgaaatgt	gtctctcata	ccttctttat	1440
	tacccaagaa	ttaatcttac	tcgatgtgca	agtattccag	acattatgga	acaacttcag	1500
	ttcattgggg	ttaaggagat	ctacagacca	gtcacgacct	ggcctttcat	tatcaaaagt	1560
	cccaagcaat	ataaaaacct	ttctttcatg	gatgctatga	ataagtttaa	atggactaaa	1620
	aaggaaggtc	tctccttcaa	caagctggtc	ctcagcctgc	cagtgaatgt	gagatgttcc	1680
	aagacagaca	atgctgagtg	gtcgattcca	aggaatgaca	gcattacctc	cagatataga	1740
	aagaccctat	aaagcagaac	cctttggtgt	gtggcacgtc	ttcttcctct	tccctgcaca	1800
	gagatttctc	ccatcaactt	gcttgtttgc	cttctgctac	tcagctgcac	gctgagcacc	1860
	aagagcttgt	gatcaaaatt	ctgttggact	tgacaatgtt	ttctatgatc	tgaacctgtc	1920
	atttgaagta	caggttaaag	actgtgtcca	ctttgggcat	gaagagtgtg	gagacttttc	1980
	ttccccattt	tccctccctc	ctttttcctt	tccatgttac	atgagagaca	tcaatcaggt	2040
	tetettetet	ttcttagaaa	tatctgatgt	tatatataca	tggtcaataa	aataaaactg	2100
				tgggctgtca	tgtgggaata	aaagaattct	2160
	ttctttccta	aaaaaaaaa	aaaaaaaa				2189
<	210> 117						
	211> 1763						
	212> DNA						
	213> Homo s	apiens					
<	400> 117						
	ggtggcctag	agatgctgct	gccgcggttg	cagttgtcgc	gcacgcctct	gcccgccagc	60
	ccgctccacc	gccgtagcgc	ccgagtgtcg	gggggcgcac	ccgagtcggg	ccatgaggcc	120
				cgtgctgctg			180
				ccggggaggg			240
				actgaacttt			300
				tcgagttctg			360
				gatggtgayt			420
				gccaggacct			480
				atgagccgtc			540
				ctggcatcgg			600
				atttcatttg			660
	cagcagttcc	ttctagagaa	gctgaaggtg	aggaaacaga	gctgacaaca	cctgtacttc	720
	cagaagaaac	acaggaagaa	gatgccaaaa	aaacatttaa	agaaagtaga	gaagctgcct	780
	tgaatctggc	ctacatccta	atccccagca	ttccccttct	cctcctcctt	gkggtcacca	840
	cagttgtatg	ttgggtttgg	atctgtagaa	aaagaaaacg	ggagcagcca	gaccctagca	900
	caaagaagca	acacaccatc	tggccctctc	ctcaccaggg	aaacagcccg	gacctagagg	960
	tctacaatgt	catwagaaaa	caaagcgaag	ctgacttagc	tgagacccgg	ccagacctga	1020
	agaatatttc	attccgagtg	tgttcgggag	aagccactcc	cgatgacatg	tcttgtgact	1080
	atgacaacat	ggctgtgaac	ccatcagaaa	gtgggtttgt	gactctggtg	agcgtggaga	1140
	grggatttgt	gaccaatgac	atttatgagt	tctccccaga	ccaaatgggg	aggagtaagg	1200
	agtctggatg	ggtggaaaat	gaaatatatg	gttattagga	catataaaaa	actgaaactg	1260
	acaacaatgg	aaaagaaatg	ataagcaaaa	tcctcttatt	ttctataagg	aaaatacaca	1320
	gaaggtctat	gaacaagctt	agatcaggtc	ctgtggatga	gcatgtggtc	cccacgacct	1380
				ctttatccca			1440
	ttatgagaag			acatagtaga			1500 1560

gttggttgta tctaacttt aagggacaga gctttacctg gcagtgataa agatgggctg tggagcttgg aaaaccacct ctgttttcct tgctctatac agcagcacat attatcatac

agacagaaaa tccagaatct tttcaaagcc cacatatggt agcacaggtt ggcctgtgca

WO 00/06698 PCT/US99/17130

79

```
tcggcaattc tcatatctgt ttttttcaaa gaataaaatc aaataaagag caggaaaaaa
                                                                       1740
 aaaaaaaaa aaaaaaaactc gag
                                                                       1763
<210> 118
<211> 1375
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (18)
<223> n equals a,t,g, or c
<400> 118
 ggcgggcggc gcggaagngg cggckgcgcg gccggggcag ccatgtcgcc attgtctgcg
                                                                         60
 gcgcgggcgg ccctgcgggt ctacgcggta ggcgccgcgg tgatcctggc gcagctgctg
                                                                        120
 cggcgctgcc gsgggggctt cctggagcca gttytccccc cacgacctga ccgtgtcgct
                                                                        180
 atagtgacgg gagggacaga tggcattggc tattctacag cgaacatctg gcgagacttg
                                                                        240
 gcatgcatgt tatcatagct ggaaataatg acagcaaagc caaacaagtt gtaagcaaaa
                                                                        300
 taaaagaaga aaccttgaac gacaaagtgg aatttttata ctgtgacttg gcttccatga
                                                                        360
 cttccatccg gcagtttgtg cagaagttca agatgaagaa gattcctctc catgtcctga
                                                                        420
 tcaacaatgc tggggtgatg atggtccctc agaggaaaac cagagatgga ttcgaagaac
                                                                        480
 atttcggcct gaactaccta gggcacttcc tgctgaccaa ccttctcttg gatacgctga
                                                                        540
 aagagtctgg gtcccctggc cacagtgcga gggtggtcac cgtctcctct gccacccatt
                                                                       600
 acgtcgctga gctgaacatg gatgaccttc agagcagtgc ctgctactca ccccacgcag
                                                                       660
 cctacgccca gagcaagctg gcccttgtcc tgttcaccta ccacctccag cggctgctgg
                                                                       720
 cggctgaggg aagccacgtg accgccaacg tggtggaccc cggggtggtc aacacggacs
                                                                       780
 tctacaagca cgtgttctgg gccacccgtc tggcgaagaa gcttctcggc tggttgcttt
                                                                       840
 tcaagacccc cgatgaagga gcgtggactt ccatctacgc agcagtcacc ccagagctgg
                                                                       900
 aaggagttgg tggccgttac ctatacaacg agaaagagac caagtccctc cacgtcacct
                                                                       960
 acaaccagaa actgcagcag cagctgtggt ctaagagttg tgagatgact ggggtccttg
                                                                      1020
 atgtgaccct gtgatatcct gtctcaggat agctgctgcc ccaagaaaca cattgcacct
                                                                      1080
 gccaatagct tgtgggtctg tgaagactgc ggtgtttgag tttctcacac ccacctgccc
                                                                      1140
 acagggetet gteetetagt tttgagacag etgeeteaac etetgeagaa etteaagaag
                                                                      1200
 ccaaataaac attttggagg ataatcaccc caagtggtct tcaaccataa actttgtgat
                                                                      1260
 tccaaagtgc ccagttgtca caggtgccat aaataattac attttccaac ataaaaaaaa
                                                                      1320
 1375
<210> 119
<211> 1022
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (937)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (990)
<223> n equals a,t,g, or c
<400> 119
ggcacgagag agccrgcggc ccgagccggg gccatggcga actgctgagc tgcgtcctag
                                                                        60
gcccccggct ctacaaaatc taccgggaga gggactctga aagggccccg gccagcgtcc
                                                                       120
ctgagacgcc aacggcagtc actgcccccc attccagctc ctgggatacg tactatcagc
                                                                       180
```

```
cccgtgccct ggagaaacat gctgacagca tcctggcact ggcttcagta ttctggtcca
                                                                         240
 tetettatta eteeteece tregeettet tetaettgta caggaaaggt taettgagtt
                                                                         300
 tgtccaaagt ggtgccgttt tctcactatg ctgggacatt gctgctactt ctggcaggtg
                                                                         360
 tggcctgctc cgaggcattg gccgctggac caacccccag taccggcagt tcatcaccat
                                                                         420
 cttggaagca acacatcgga accagtette agaaaacaag aggeagettg ccaactacaa
                                                                         480
 ctttgacttc cggagctggc cagtcgactt ccactgggaa gaacccagca gccggaagga
                                                                         540
 gtctcgaggg ggcccttccc gccggggtgt ggccctgctt cgcccagagc ccctgcaccg
                                                                         600
 ggggacagca gacaccctcc tcaaccgggt taagaagctg ccttgtcaga tcaccagcta
                                                                         660
 cctggtggcg cacaccctag ggcgccggat gctgtatcca ggcţctgtgt acctgctgca
                                                                         720
 gaaggccctc atgcctgtgc tgctgcaggg ccaggcccga ctggtggaag agtgtaatgg
                                                                         780
 gcgccgggca aagctgctgg cctgtgatgg caatgagatt gacaccatgt ttgtggaccg
                                                                         840
 gcgggggaca gctgagcccc aggacagaag ctggtgatct gctgtgargg gaatgctggg
                                                                         900
 ttttatgarg tgggctgcgt ctccamgccc ctggaanctg gatattcatc ckgggtggaa
                                                                         960
 tcatccagct ttgctggaac acagggggtn ccattcccgc aaaatgaagg ctaatgccat
                                                                        1020
                                                                        1022
<210> 120
<211> 2311
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (654)
<223> n equals a,t,g, or c
<220>
<221> SITE
<222> (2293)
<223> n equals a,t,g, or c
<400> 120
 togacccacg cgtccgcttc ctcatgctgc ccgctcgcct accgttcagg ctgctgagcc
                                                                         60
 ttttccttcg tggatccgct cccacggcag cgcgccatgg cctccgggag ccgctcctgg
                                                                        120
agaggaggtg cgctgccgcc tcctccttcc agcactcatc gagtctggga cgcgagcttc
                                                                        180
cttatgaccc cgtggacacg gagggetttg gagaaggtgg tgacatgcag gagcgttttc
                                                                        240
 tgttcccgga gtacatcctg gatccggagc cgcaacccac ccgcgaaaag cagctgcagg
                                                                        300
agctccagca acagcaggag gaggaggagc gacagaggca gcagcggcgg gaggagcggc
                                                                        360
gacagcaaaa cctacgggcc aggtcccggg agcacccggt cgtggggcac ccggacccgg
                                                                        420
cattgccgcc cagcggcgtg aactgctcgg gctgcggggc asasctgcac tgccaggacg
                                                                        480
ccggagtgcc cggctacctg ccccgagaga agttcctccg cacggcggag gcagacggcg
                                                                        540
ggctggcacg gaccgtgtgc cagcgctgct ggctgctgtc gcaccaccgg cgcgctctac
                                                                        600
gcctgcaggt gagccgcgag cagtacctgg agctggtgag cgccgcgttg cggnggcccg
                                                                        660
gcccctccct ggtgctctac atggtggacc tgctggacct gcccgacgcc ctgctgcccg
                                                                        720
acttgcccgc gctggtgggc cccaagcagc tgatcgtgct gggaaacaaa gtggacctcc
                                                                        780
tgccccagga tgctcctggc taccggcaga ggctgcggga gcgactgtgg gaggactgtg
                                                                        840
cccgcgccgg gctcctgctg gcccctggca ccaagggcca cagcgccccg tcaaggacga
                                                                        900
gccacaggac ggggagaatc cgaatccgcc gaactggtcc cgcacagtgg tcagggacgt
                                                                        960
gcggctgatc agcgccaaga ccggctatgg agtggaagag ttgatctctg cccttcagcg
                                                                       1020
ctcctggcgc taccgtgggg acgtctactt agtgggcgcc accaacgccg gcaaatccac
                                                                       1080
tetetttaac aegeteetgg agteegatta etgeaetgee aagggeteeg aggeeatega
                                                                       1140
cagagccacc atctcccctt ggccaggtac tacattaaac cttctgaagt ttcctatttg
                                                                       1200
caacccaact ccttacagaa tgtttaaaag gcatcaaaga cttaaaaaag attcaactca
                                                                       1260
agctgaagaa gatcttagtg agcaagaaca aaatcagctt aatgtcctca aaaagcatgg
                                                                       1320
ttatgtcgta ggaagagttg gaaggacatt cttgtattca gaagaacaga aggataacat
                                                                       1380
tccctttgag tttgatgctg attcacttgc ctttgacatg gaaaatgacc ctgttatggg
                                                                       1440
tacacacaaa tccaccaaac aagtagaatt gactgcacaa gatgtgaaag atgcccactg
                                                                       1500
gttttatgac acccctggaa ttacaaaaga aaattgtatt ttaaatcttc taacagaaaa
                                                                       1560
```

```
agaagtaaat attgttttgc caacacagtc cattgttcca agaacttttg tgcttaaacc
                                                                       1620
  aggaatggtt ctgtttttgg gtgctatagg ccgcatagat ttcctgcagg gaaatcagtc
                                                                        1680
  agcttggttt acagtcgtgg cttccaacat cctccctgtg catatcacct ccttggacag
                                                                       1740
  ggcagacgct ctgtatcaga agcatgcagg tcatacgtta ctccagattc caatgggtgg
                                                                       1800
  aaaagaacga atggcrggat ttcctcctct tgttgctgaa gacattatgt taaaagaagg
                                                                       1860
  actgggggca tctgaagcag tggccgacat caagttttcc tctgcaggtt gggtttcagt
                                                                       1920
  aacacctaat tttaaggaca gactgcatct ccgaggctat acacctgaag gaacagtttt
                                                                       1980
 gaccgtccgg ccccctctct tgccatatat tgttaacatc aaaggacagc gcatcaagaa
                                                                       2040
 aagtgtggcc tataaaacca agaagcctcc ttcccttatg tacaacgtga ggaagaagaa
                                                                       2100
 aggaaagata aatgtatgag accgaccttg ttcactccag atattaactg tattgaacac
                                                                       2160
 aacaaaatac attgaatttg tattaaacat ataacgcata aataaagctc ccattcttac
                                                                       2220
 ccttaaaaaa aaaaaaaag ggcggccgct ctagaggatc caagcttacg tacgcgtgca
                                                                       2280
  tgcgacgtca tanctcgtct ataggaactg g
                                                                       2311
<210> 121
<211> 1286
<212> DNA
<213> Homo sapiens
<220>
<221> SITE
<222> (1284)
<223> n equals a,t,g, or c
<400> 121
 gggcgcgcgg gtgaaaggcg cattgatgca gcctgcggcg gcctcggagc gcggcggagc
                                                                         60
 agacgctgac cacgttcctc tcctcggtct cctccgcctc cagctccgcg ctgcccggca
                                                                        120
 geogggagee atgegaeece agggeeecge egecteeeg cageggetee geggeeteet
                                                                        180
 gctgctcctg ctgctgcagc tgcccgcgcc gtcgagcgcc tctgagatcc ccaaggggaa
                                                                       240
 gcaaaaggcg cacteeggca gagggaggtg gtggacetgt ataatggaat gtgettacaa
                                                                       300
 gggccagcag gagtgcctgg tcgagacggg agccctgggg ccaatggcat tccgggtaca
                                                                       360
 cctgggatcc caggtcggga tggattcaaa ggagaaaagg gggaatgtct gagggaaagc
                                                                       420
 tttgaggagt cctggacacc caactacaag cagtgttcat ggagttcatt gaattatggc
                                                                       480
 atagatettg ggaaaattge ggagtgtaca tttacaaaga tgegtteaaa tagtgeteta
                                                                       540
 agagttttgt tcagtggctc acttcggcta aaatgcagaa atgcatgctg tcagcgttgg
                                                                       600
 tatttcacat tcaatggagc tgaatgttca ggacctcttc ccattgaagc tataatttat
                                                                       660
 ttggaccaag gaagccctga aatgaattca acaattaata ttcatcgcac ttcttctgtg
                                                                       720
 gaaggacttt gtgaaggaat tggtgctgga ttagtggatg ttgctatctg ggttggcact
                                                                       780
 tgttcagatt acccaaaagg agatgcttct actggatgga attcagtttc tcgcatcatt
                                                                       840
 attgaagaac taccaaaata aatgctttaa ttttcatttg ctacctcttt ttttattatg
                                                                       900
 ccttggaatg gttcacttaa atgacatttt aaataagttt atgtatacat ctgaatgaaa
                                                                       960
 agcaaagcta aatatgttta cagaccaaag tgtgatttca cactgttttt aaatctagca
                                                                      1020
 ttattcattt tgcttcaatc aaaagtggtt tcaatatttt ttttagttgg ttagaatact
                                                                      1080
 ttottoatag toacattoto toaacotata atttggaata ttgttgtggt ottttgtttt
                                                                      1140
 ttctcttagt atagcatttt taaaaaaata taaaagctac caatctttgt acaatttgta
                                                                      1200
 1260
 aaaaaaaaa aaaaaaaaa aaanaa
                                                                      1286
<210> 122
<211> 1380
<212> DNA
<213> Homo sapiens
<400> 122
cagacccgcg gggcaaacgg actggggcca agagccggga gcgcgggcgc aaaggcacca
                                                                        60
gggcccgccc agggcgccgc gcacacggcc ttgggggttc tgcgggcctt cgggtgcgcg
                                                                       120
totogoctot agocatgggg toogoagogt tggagatoot gggcotggtg otgtgootgg
                                                                       180
```

```
tgggctgggg gggtctgatc ctggcgtgcg ggctgcccat gtggcaggtg accgccttcc
                                                                         240
 tggaccacaa catcgtgacg gcgcagacca cctggaaggg gctgtggatg tcgtgcgtgg
                                                                         300
 tgcagagcac gggcacatgc agtgcaaagt gtacgactcg gtgctggctc tgagcaccga
                                                                         360
 ggtgcaggcg gcgcgggcgc tcaccgtgag cgccgtgctg ctggcgttcg ttgcgctctt
                                                                         420
 cgtgaccctg gcgggcgcgc agtgcaccac ctgcgtggcc ccgggcccgg ccaaggcgcg
                                                                         480
 tgtggccctc acgggaggcg tgctctacct gttttgcggg ctgctggcgc tcgtgccact
                                                                         540
 ctgctggttc gccaacattg tcgtccgcga gttttacgac ccgtctgtgc ccgtgtcgca
                                                                         600
 gaagtacgag ctgggcgcac gctgtacatc ggctgggcgg ccaccgcgct gctcatggta
                                                                         660
 ggcggctgcc tcttgtgctg cggcgcctgg gtctgcaccg gccgtcccga cctcagcttc
                                                                         720
 cccgtgaagt actcagegee geggeggeee aeggeeaceg gegaetaega caagaagaae
                                                                         780
 tacgtctgag ggcgctgggc acggccgggc ccctcctgcc agccacgcct gcgaggcgtt
                                                                         840
 ggataagcct ggggagcccc gcatggaccg cggcttccgc cgggtagcgc ggcgcgcagg
                                                                         900
 ctcctcggaa cgtccggctc tgcgccccga cgcggctcct ggatccgctc ctgcctgcgc
                                                                         960
 ccgcagctga ccttctcctg ccactagccc ggccctgccc ttaacagacg gaatgaagtt
                                                                        1020
 tccttttctg tgcgcggcgc tgtttccata ggcagagcgg gtgtcagact gaggatttcg
                                                                        1080
 cttcccctcc aagacgctgg gggtcttggc tgctgcctta cttcccagag gctcctgctg
                                                                        1140
 acttcggagg ggcggatgca gagcccaggg cccccaccgg aagatgtgta cacctggtct
                                                                        1200
 ttactccatc ggcagggccc gagcccaggg accagtgact tggcctggac ctcccggtct
                                                                        1260
 cactccagca tetececagg caaggettgt gggcacegga gettgagaga gggegggagt
                                                                        1320
 gggaaggcta agaatctgct tagtaaatgg tttgaactct caaaaaaaaa aaaaaaaaa
                                                                       1380
<210> 123
<211> 3793
<212> DNA
<213> Homo sapiens
```

<220> <221> SITE <222> (1102) <223> n equals a,t,g, or c<220> <221> SITE <222> (1132) <223> n equals a,t,g, or c <220> <221> SITE <222> (1199) <223> n equals a,t,g, or c <220> <221> SITE <222> (1228) <223> n equals a,t,g, or c<220> <221> SITE <222> (1229) <223> n equals a,t,g, or c <220> <221> SITE <222> (1231) <223> n equals a,t,g, or c <220>

<221> SITE

<222> (3176) <223> n equals a,t,g, or c

```
<400> 123
ggggcttcat acaggaaatc tattgctgtg tcaagttcca gagaaaagct tctgttcgtc
                                                                          60
caagttacta accaggctaa accacataga cgtgaaggaa ggggctagaa ggaagggagt
                                                                         120
gccccactgt tgatggggta agaggatcct gtactgagaa gttgaccaga gagggtctca
                                                                         180
ccatgcgcac agttccttct gtaccagtgt ggaggaaaag tactgagtga agggcagaaa
                                                                         240
aagagaaaac agaaatgctc tgcccttgga gaactgctaa cctagggcta ctgttgattt
                                                                         300
tgactatctt cttagtggcc gaagcggagg gtgctgctca accaaacaac tcattaatgc
                                                                         360
tgcaaactag caaggagaat catgctttag cttcaagcag tttatgtatg gatgaaaaac
                                                                         420
agattacaca gaactactcg aaagtactcg cagaagttaa cacttcatgg cctgtaaaga
                                                                         480
tggctacaaa tgctgtgctt tgttgccctc ctatcgcatt aagaaatttg atcataataa
                                                                         540
catgggaaat aatcctgaga ggccagcctt cctgcacaaa agcctacaag aaagaaacaa
                                                                         600
atgagaccaa ggaaaccaac tgtactgatg agagaataac ctgggtctcc agacctgatc
                                                                        660
agaattcgga ccttcagatt cgtaccgtgg ccatcactca tgacgggtat tacagatgca
                                                                        720
taatggtaac acctgatggg aatttccatc gtggatatca cctccaagtg ttagttacac
                                                                        780
ctgaagtgac cctgtttcaa aacaggaata gaactgcagt atgcaaggca gttgcaggga
                                                                        840
agccagctgc gcatatctcc tggatcccag agggcgattg tgccactaag caagaatact
                                                                        900
ggagcaatgg cacagtgact gttaagagta catgccactg ggaggtccac aatgtgtcta
                                                                        960
ccgtgaactg ccacgtctcc catttgactg gcaacaagag tctgtacata gagctacttc
                                                                       1020
ctgttccagg tgccaaaaaa tcatcaaaat tatatattcc atatatcatc cttactatta
                                                                       1080
ttattttgac catcgtggga tncatttggt tgttgaaagt caatggctgc anaaaatata
                                                                       1140
aattgaataa accagaatct actccagttg ttgaggagga tgaaatgcag ccctatgcnt
                                                                       1200
tttacacaga gaagaacaat cctctctnng ntactacaaa caaggtgaag gcatctgagg
                                                                       1260
cattacaaag tgaagttgac acagacctcc atactttata agttgttgga ctctagtacc
                                                                       1320
aagaaacaac aacaaacgag atacattata attactgtct gattttctta cagttctaga
                                                                       1380
atgaagactt atattgaaat taggttttcc aaggttctta gaagacattt taatggattc
                                                                       1440
tcattcatac ccttgtataa ttggaatttt tgattcttag ctgctaccag ctagttctct
                                                                       1500
gaagaactga tgttattaca aagaaaatac atgcccatga ccaaatattc aaattgtgca
                                                                       1560
ggacagtaaa taatgaaaac caaatttcct caagaaataa ctgaagaagg agcaagtgtg
                                                                       1620
aacagtttct tgtgtatcct ttcagaatat tttaatgtac atatgacatg tgtatatgcc
                                                                       1680
tatggtatat gtgtcaattt atgtgtcccc ttacatatac catgcaccct atctttgtca
                                                                       1740
aggcaccagt gggaacaata cactgcatta ctgttctata catatgaaaa cctaataata
                                                                       1800
taagtottag agatcatttt atatcatgac aagtagagct acctcattct ttttaatggt
                                                                       1860
tatataaaat tooattgtat agttatatca ttatttaatt aaaaacaacc ctaatgatgg
                                                                       1920
atatttagat tcttttaagt tttgtttatt tcttttaagt tttgtttgtg gtataaacaa
                                                                       1980
taccacatag aatgittett gigcatatat etetitgitt tigagtatat eigtaggata
                                                                       2040
actttcttga gtggaattgt caggtcaaag ggtttgtgca ttttactatt gatatatatg
                                                                       2100
ttaaattgtg tcaaatatat atgtcaaatt ccctccaaca ttgtttaaat gtgcctttcc
                                                                       2160
ctaaatttct attttaataa ctgtactatt cctgcttcta cagttgccac tttctctttt
                                                                       2220
taatcaacca gattaaatat gatgtgagat tataataaga attatactat ttaataaaaa
                                                                       2280
tggatttata tttttggtca tgtttgtaag agagtgaatg cacgtgtgag aacattagct
                                                                       2340
tcttctgaac tcattatatc tccacagagg tgttgatact tgatgcctaa cagttttgca
                                                                       2400
gatgtgctac attggaattg tgtattttta tggtgtacat tctattgtga tatatttatt
                                                                       2460
gaataattaa tgtctattga ccatataagt ggcgaaaaat gcaccataga ggacatgggg
                                                                       2520
tatttattta caaactatga gctacataat aagcaagtgg ccatgggatg gcatgaccct
                                                                       2580
cccctccata tttttgtgga gcaaaatatt ggcaatgttt atgtaaatca ttgttaatat
                                                                       2640
catgaaatta tttttaatta aaaacataag tctatttgct ccatagcaga aaaaacatga
                                                                       2700
gaagtttttt catcatgata gaaattgaaa caaaatatat tcattcttca atcataccat
                                                                       2760
ctgagatttt taagacagct attttgtctt ataagtatat ttttctccct ctagacattt
                                                                      2820
cagttactat ggattttgtc ctcaaaggga cytttagtct aattttggga tgtaaagcta
                                                                      2880
atcttaatga cacttggcac atgatatttt gatcaagcca ttttgacttg accaaaaagc
                                                                      2940
agtgtccatt aggtttctgc atataaatat taccaagcaa tgttcacaat agacatcatt
                                                                      3000
acactgteet tgaaatttat taattettea tecaaccetg gttgagetga ggeteatagt
                                                                      3060
taggttcaag actatctgtt taaatattac tgaaaaacaa agtaagacag tactatgctt
                                                                      3120
acctettaae ttgataatgt caaaccagge atgttaaatg acateataga aaaganttea
                                                                      3180
agataattta tagaagttaa attatattgt acaqaaaata attgtatgaa aatctctact
                                                                      3240
atggggctgg aacatggttg aacattagaa tgatataaaa aattatatat attctccaaa
                                                                      3300
```

tccacgctag acctgtcaaa ttagagaatc tagagattag acctggcgtg tcagcaaggt catccaggaa gcagaggctg agacggagtt aggtgtgatt acttacatag tcgattacat tttacaaata acatttata tgtctcattt actgtgcttt ctccccatcc cattttgtat cattactcta ataatttgaa agtaattcat tttctctctc tttctgtctc tctctctttc aatatctcta ataatttgaa cggctttact gagaatataat tgataattt aagtgtacag tttgttaaat tttgcacata tttgcacata tttacaaatgt ggactttggt aaatgttgac atagtttac aaaaaaaaaa
<210> 124 <211> 370 <212> PRT <213> Homo sapiens
<220> <221> SITE <222> (370) <223> Xaa equals stop translation
<pre><400> 124 Met Leu Gly Ala Phe Val Trp Pro Ser Leu Leu Leu Leu Ala Ala 1 5 10 15</pre>
Cys Ile Cys Leu Leu Thr Phe Ile Asn Cys Ala Tyr Val Lys Trp Gly 20 25 30
Thr Leu Val Gln Asp Ile Phe Thr Tyr Ala Lys Val Leu Ala Leu Ile 35 40 45
Ala Val Ile Val Ala Gly Ile Val Arg Leu Gly Gln Gly Ala Ser Thr 50 55 60
His Phe Glu Asn Ser Phe Glu Gly Ser Ser Phe Ala Val Gly Asp Ile 65 70 75 80
Ala Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr Ser Gly Trp Asp Thr 85 90 95
Leu Asn Tyr Val Thr Glu Glu Ile Lys Asn Pro Glu Arg Asn Leu Pro 100 105 110
Leu Ser Ile Gly Ile Ser Met Pro Ile Val Thr Ile Ile Tyr Ile Leu 115 120 125
Thr Asn Val Ala Tyr Tyr Thr Val Leu Asp Met Arg Asp Ile Leu Ala 130 135 140
Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln Ile Phe Gly Ile Phe 145 150 155 160
Asn Trp Ile Ile Pro Leu Ser Val Ala Leu Ser Cys Phe Gly Gly Leu 165 170 175
Asn Ala Ser Ile Val Ala Ala Ser Arg Leu Phe Phe Val Gly Ser Arg 180 185 190

Glu Gly His Leu Pro Asp Ala Ile Cys Met Ile His Val Glu Arg Phe

Thr Pro Val Pro Ser Leu Leu Phe Asn Gly Ile Met Ala Leu Ile Tyr 210 220

Leu Cys Val Glu Asp Ile Phe Gln Leu Ile Asn Tyr Tyr Ser Phe Ser 225 230 235 240

Tyr Trp Phe Phe Val Gly Leu Ser Ile Val Gly Gln Leu Tyr Leu Arg 245 250 255

Trp Lys Glu Pro Asp Arg Pro Arg Pro Leu Lys Leu Ser Val Phe Phe 260 265 270

Pro Ile Val Phe Cys Leu Cys Thr Ile Phe Leu Val Ala Val Pro Leu 275 280 285

Tyr Ser Asp Thr Ile Asn Ser Leu Ile Gly Ile Ala Ile Ala Leu Ser 290 295 300

Gly Leu Pro Phe Tyr Phe Leu Ile Ile Arg Val Pro Glu His Lys Arg 305 310 315 320

Pro Leu Tyr Leu Arg Arg Ser Trp Gly Leu Pro Gln Gly Thr Ser Arg 325 330 335

Ser Cys Val Cys Gln Leu Leu Gln Lys Trp Ile Trp Lys Met Glu Glu 340 345 350

Arg Cys Pro Ser Asn Gly Ile Pro Ser Leu Thr Lys His His Leu Glu 355

Ser Xaa 370

<210> 125

<211> 86

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (86)

<223> Xaa equals stop translation

<400> 125

Met Gly Phe Trp Cys Gly Cys Pro Phe Cys Leu Leu Val Val Leu Leu 1 5 10 15

Thr Asp Arg Thr Leu Ser Cys Arg Ser Val Gly Val Pro Cys Asn Val

Arg Cys Gln Cys Ala Pro Ala Gly Gly Cys Leu Pro Val Arg Leu Leu 35 40

Ala Gly Gln Gly Ser Gly Thr His Leu Arg Arg Gln Ser Ala Arg Ser 50 60

```
Gln Ile Ser Ser Cys Met Leu Gly Glu Pro Leu Leu Ser Ser Lys Leu
                      70
Ser Asp Arg Asp Ile Xaa
<210> 126
<211> 44
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (44)
<223> Xaa equals stop translation
<400> 126
Met Tyr Thr Lys Thr His Lys Phe Lys Phe Tyr Asn Phe Leu Ser Leu
```

Trp Ile Trp Lys Ile Phe Phe Leu Leu Phe Phe Ile Leu Ile Val Ala

10

Leu Ala Phe Pro Ile Pro Cys Leu Ser Ile Phe Xaa 40

<210> 127 <211> 319 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (264) <223> Xaa equals any of the naturally occurring L-amino acids <220> <221> SITE <222> (303) <223> Xaa equals any of the naturally occurring L-amino acids <400> 127 Met Asn Thr Asp His Leu Arg Leu Thr Val Pro Asn Gly Ile Gly Ala Leu Lys Leu Arg Glu Met Glu His Tyr Phe Ser Gln Gly Leu Ser Val 20 25 Gln Leu Phe Asn Asp Gly Ser Lys Gly Lys Leu Asn His Leu Cys Gly Ala Asp Phe Val Lys Ser His Gln Lys Pro Pro Gln Gly Met Glu Ile

Lys Ser Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile

Val Tyr Tyr His Asp Ala Asp Gly His Phe His Leu Ile Asp Gly
85 90 95

Asp Lys Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu Leu Val 100 \$105\$

Glu Ile Gly Glu Ser Leu Asn Ile Gly Val Val Gln Thr Ala Tyr Ala 115 120 125

Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu Val Met Lys Val Pro Val 130 140

Tyr Cys Thr Lys Thr Gly Val Lys His Leu His His Lys Ala Gln Glu 145 150 155 160

Phe Asp Ile Gly Val Tyr Phe Glu Ala Asn Gly His Gly Thr Ala Leu 165 170 175

Phe Ser Thr Ala Val Glu Met Lys Ile Lys Gln Ser Ala Glu Gln Leu 180 185 190

Glu Asp Lys Lys Arg Lys Ala Ala Lys Met Leu Glu Asn Ile Ile Asp 195 200 205

Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile Ser Asp Met Leu Val Ile 210 215 220

Glu Ala Ile Leu Ala Leu Lys Gly Leu Thr Val Gln Gln Trp Asp Ala 225 230 235 240

Leu Tyr Thr Asp Leu Pro Asn Arg Gln Leu Lys Val Gln Val Ala Asp 245 250 255

Arg Arg Val Ile Ser Thr Thr Xaa Ala Glu Arg Gln Ala Val Thr Pro 260 265 270

Pro Gly Leu Gln Glu Ala Ile Asn Asp Leu Val Lys Lys Tyr Lys Leu 275 280 285

Ser Arg Ala Phe Val Arg Pro Ser Gly Thr Glu Asp Val Val Xaa Ser 290 295 300

Ile Cys Arg Ser Arg Leu Thr Arg Lys Cys Arg Ser Pro Cys Thr 305 315

<210>.128

<211> 46

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (46)

<223> Xaa equals stop translation

<400> 128

Met Asp Met Val Cys Phe Cys Ile Tyr Leu Gly Leu Leu Lys Phe Ile

1 10 15

```
Ser Ala Ile Phe Cys Ser Phe Ser Glu Glu Val Leu Tyr Ile Ser Phe
                                   25
  Val Lys Cys Ile Pro Lys Tyr Phe Val Glu Met Leu Leu Xaa
           35
                               40
  <210> 129
  <211> 709
  <212> PRT
  <213> Homo sapiens
 <220>
 <221> SITE
 <222> (189)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (275)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (414)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (438)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
· <222> (641)
 <223> Xaa equals any of the naturally occurring L-amino acids
<220>
 <221> SITE
<222> (643)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (696)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (697)
<223> Xaa equals any of the naturally occurring L-amino acids
<400> 129
Met Ala Gly Leu Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu
Leu Leu Gly Ala Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile
```

- Ala Leu Pro Arg Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr
 35 40 45
- Pro Thr Leu Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser Lys Arg His 50 60
- Ile Thr Met Leu Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe 65 70 75 80
- Ser Cys Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn 85 90 95
- Ile Asp Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln
 100 105 110
- Pro Ser Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp 115 120 125
- Val Lys Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe Ser Ile Pro 130 135 140
- Arg Leu Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp Gly Val Thr 145 150 155 160
- His Ser Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg Ile Gly Thr 165 170 175
- Phe Cys Ser Asn Gly Thr Val Ser Arg Ile Lys Met Xaa Glu Gly Val 180 185 190
- Lys Met Ala Leu His Leu Pro Trp Phe His Pro Arg Asn Val Ser Gly 195 200 205
- Phe Ser Ile Ala Asn Arg Ser Ser Ile Lys Arg Leu Cys Ile Ile Glu 210 215 220
- Ser Val Phe Glu Gly Glu Gly Ser Ala Thr Leu Met Ser Ala Asn Tyr 225 230 235 240
- Pro Glu Gly Phe Pro Glu Asp Glu Leu Met Thr Trp Gln Phe Val Val 245 250 255
- Pro Ala His Leu Arg Ala Ser Val Ser Phe Leu Asn Phe Asn Leu Ser 260 265 270
- Asn Cys Xaa Arg Lys Glu Glu Arg Val Glu Tyr Tyr Ile Pro Gly Ser 275 280 285
- Thr Thr Asn Pro Glu Val Phe Lys Leu Glu Asp Lys Gln Pro Gly Asn 290 295 300
- Met Ala Gly Asn Phe Asn Leu Ser Leu Gln Gly Cys Asp Gln Asp Ala 305 310 315 320
- Gln Ser Pro Gly Ile Leu Arg Leu Gln Phe Gln Val Leu Val Gln His 325 330 335
- Pro Gln Asn Glu Ser Asn Lys Ile Tyr Val Val Asp Leu Ser Asn Glu

				34	0					3	45						350)		
A	rg A	la	Met 355	Se	r Le	eu T	hr I	le	Gl: 360	1 P:	ro i	Arg	Pro	o Va	1 L 3	ys 65	Glr	ı Se	er	Arg
Ly	ys P 3	he 70	Val	Pr	o G1	.у С	ys F 3	he 75	Val	. C)	ys I	eu	Glu	1 Se 38	r A O	rg	Thr	. G?	's	Ser
Se 38	er A	sn :	Leu	Thi	r Le	u Tl 39	nr S 90	er	Gly	S€	er L	ys	His 395	Ly	s I	le	Ser	Ph		Leu 400
СУ	's As	gp /	Asp	Leu	Th 40	r Ai 5	g L	eu '	Trp	Me	t A 4	sn 10	Val	Gli	u Ly	/s :	Xaa	I1 41		Ser
Су	s Th	ır A	qz	His 420	Ar	д Ту	r C	ys (Gln	Ar 42	g L 5	ys _.	Ser	Туз	s Se	er I	Leu 130	Gl:	n '	Val
	o Se	_						4	14 U						44	5				
	s Le 45						4.	5						460						
						- T	,					4	175						4	80
	Ty				403						49	0						495		
	Phe								:	202						5:	10			
	Se ₁		. •					34	20						525	i				
	Arg 530						25.	,						540						
	Val					330						5	55						56	0
Thr	Pro	As	n T	rp :	Asp 565	Arg	Gly	Le	u P	ro	Ser 570	Le	eu T	hr	Ser	۷a	1 S	er 75	Tr	p
Asn	Ile	Se	r V 5	al 1 80	Pro	Arg	Asp	Gl	n V 5	al 85	Ala	C.	s I	eu '	Thr	Ph 59		he	Ly	s
Glu	Arg	Se:	r G:	ly (/al	Val	Cys	G1: 60:	n T	hr	Gly	Ar	g A	la 1	Phe 605	Me	t I	le	I 1	e
Gln	Glu 610	Glr	ı Aı	rg I	hr .	Arg	Ala 615	Glı	u G	lu	Ile	Ph	e S 6	er 1 20	Leu	Ası	9 G	lu .	Ası	þ
Val 625	Leu	Pro	, L	/s P	ro :	Ser 630	Phe	His	s Hi	is :	His	Se 63	r P	he 1	rp	Va]	l A:		Ile 540	

Xaa Asn Xaa Ser Pro Thr Ser Gly Lys Gln Leu Asp Leu Leu Phe Ser 645 650 655

Val Thr Leu Thr Pro Arg Thr Val Asp Leu Thr Val Ile Leu Ile Ala 660 665 670

Ala Val Gly Gly Val Leu Leu Ser Ala Leu Gly Leu Ile Ile 675 680 685

Cys Cys Val Lys Lys Lys Xaa Xaa Thr Arg Gly Pro Ala Val Gly 690 700

Ile Tyr Asn Gly Asn 705

<210> 130

<211> 415

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (415)

<223> Xaa equals stop translation

<400> 130

Met Thr Lys Ala Arg Leu Phe Arg Leu Trp Leu Val Leu Gly Ser Val

1 5 10 15

Phe Met Ile Leu Leu Ile Ile Val Tyr Trp Asp Ser Ala Gly Ala Ala 20 25 30

His Phe Tyr Leu His Thr Ser Phe Ser Arg Pro His Thr Gly Pro Pro 35 40 45

Leu Pro Thr Pro Gly Pro Asp Arg Asp Arg Glu Leu Thr Ala Asp Ser 50 60

Asp Val Asp Glu Phe Leu Asp Lys Phe Leu Ser Ala Gly Val Lys Gln 65 70 75 80

Ser Asp Leu Pro Arg Lys Glu Thr Glu Gln Pro Pro Ala Pro Gly Ser 85 90 95

Met Glu Glu Asn Val Arg Gly Tyr Asp Trp Ser Pro Arg Asp Ala Arg
100 105 110

Arg Ser Pro Asp Gln Gly Arg Gln Gln Ala Glu Arg Arg Ser Val Leu 115 120 125

Arg Gly Phe Cys Ala Asn Ser Ser Leu Ala Phe Pro Thr Lys Glu Arg

Ala Phe Asp Asp Ile Pro Asn Ser Glu Leu Ser His Leu Ile Val Asp 145 150 150 160

Asp Arg His Gly Ala Ile Tyr Cys Tyr Val Pro Lys Val Ala Cys Thr 165 170 175

Asn Trp Lys Arg Val Met Ile Val Leu Ser Gly Ser Leu Leu His Arg 180 185 190

- Gly Ala Pro Tyr Arg Asp Pro Leu Arg Ile Pro Arg Glu His Val His 195 200 205
- Asn Ala Ser Ala His Leu Thr Phe Asn Lys Phe Trp Arg Arg Tyr Gly 210 215 220
- Lys Leu Ser Arg His Leu Met Lys Val Lys Leu Lys Lys Tyr Thr Lys 225 230 235 240
- Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe Arg 245 250 250
- Ser Lys Phe Glu Leu Glu Asn Glu Glu Phe Tyr Arg Lys Phe Ala Val 260 265 270
- Pro Met Leu Arg Leu Tyr Ala Asn His Thr Ser Leu Pro Ala Ser Ala 275 280 285
- Arg Glu Ala Phe Arg Ala Gly Leu Lys Val Ser Phe Ala Asn Phe Ile
 290 295 300
- Gln Tyr Leu Leu Asp Pro His Thr Glu Lys Leu Ala Pro Phe Asn Glu 305
- His Trp Arg Gln Val Tyr Arg Leu Cys His Pro Cys Gln Ile Asp Tyr 325 330 335
- Asp Phe Val Gly Lys Leu Glu Thr Leu Asp Glu Asp Ala Ala Gln Leu 340 345 350
- Leu Gln Leu Leu Gln Val Asp Arg Gln Leu Arg Phe Pro Pro Ser Tyr 355 360 365
- Arg Asn Arg Thr Ala Ser Ser Trp Glu Glu Asp Trp Phe Ala Lys Ile 370 380
- Pro Leu Ala Trp Arg Gln Gln Leu Tyr Lys Leu Tyr Glu Ala Asp Phe 385 390 395 400
- Val Leu Phe Gly Tyr Pro Lys Pro Glu Asn Leu Leu Arg Asp Xaa 405 410 415
- <210> 131
- <211> 242
- <212> PRT
- <213> Homo sapiens
- <400> 131
- Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys

 1 5 10 15
- Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys 20 25 30
- Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val 35 40 45

Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala 50 60

Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile
65 70 75 80

Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser 85 90 95

Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala 100 105 110

Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr 115 120 125

Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu 130 135 140

Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly 145 150 155 160

Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala 165 170 175

Leu Ala Phe Leu Ser Gly Tyr Tyr Val Thr Leu Ala Ala Gln Ile Leu 180 185 190

Ala Val Leu Leu Pro Pro Val Met Leu Leu Ile Asp Gly Asn Val Ala 195 200 205

Tyr Trp His Asn Thr Arg Arg Val Glu Phe Trp Asn Gln Met Lys Leu 210 215 220

Leu Gly Glu Ser Val Gly Ile Phe Gly Thr Ala Val Ile Leu Ala Thr 225 230 235 240

Asp Gly

<210> 132

<211> 313

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (313)

<223> Xaa equals stop translation

<400> 132

Met Glu Ser Leu Tyr Asp Leu Trp Glu Phe Tyr Leu Pro Tyr Leu Tyr

1 10 15

Ser Cys Ile Ser Leu Met Gly Cys Leu Leu Leu Leu Cys Thr Pro 20 25 30

Val Gly Leu Ser Arg Met Phe Thr Val Met Gly His Leu Leu Val Lys 35 40

- Pro Thr Ile Leu Glu Asp Leu Asp Glu Gln Ile Tyr Ile Ile Thr Leu 50 55 60
- Glu Glu Glu Ala Leu Gln Arg Arg Leu Asn Gly Leu Ser Ser Val
 65 70 75 80
- Glu Tyr Asn Ile Met Glu Leu Glu Glu Glu Leu Glu Asn Val Lys Thr
 85 90 95
- Leu Lys Thr Lys Leu Glu Arg Arg Lys Lys Ala Ser Ala Trp Glu Arg
- Asn Leu Val Tyr Pro Ala Val Met Val Leu Leu Leu Ile Glu Thr Ser
- Ile Ser Val Leu Leu Val Ala Cys Asn Ile Leu Cys Leu Leu Val Asp
- Ser Leu Ser Thr Phe Gly Phe Val Gly Ala Ala Leu Glu Ile Ile Leu 165 170 175
- Ile Phe Tyr Leu Met Val Ser Ser Val Val Gly Phe Tyr Ser Leu Arg
- Phe Phe Gly Asn Phe Thr Pro Lys Lys Asp Asp Thr Thr Met Thr Lys 195 200 205
- Ile Ile Gly Asn Cys Val Ser Ile Leu Val Leu Ser Ser Ala Leu Pro 210 215 220
- Val Met Ser Arg Thr Leu Gly Ile Thr Arg Phe Asp Leu Leu Gly Asp 235 240
- Phe Gly Arg Phe Asn Trp Leu Gly Asn Phe Tyr Ile Val Leu Ser Tyr 245 250 255
- Asn Leu Leu Phe Ala Ile Val Thr Thr Leu Cys Leu Val Arg Lys Phe 260 265 270
- Thr Ser Ala Val Arg Glu Glu Leu Phe Lys Ala Leu Gly Leu His Lys
 275 280 285
- Leu His Leu Pro Asn Thr Ser Arg Asp Ser Glu Thr Ala Lys Pro Ser 290 295 300
- Val Asn Gly His Gln Lys Ala Leu Xaa 305 310

<210> 133

<211> 183

<212> PRT

<213> Homo sapiens

<220>

```
<221> SITE
```

<222> (183)

<223> Xaa equals stop translation

<400> 133

Leu Arg Ser Tyr Met Gln Ser Val Trp Thr Glu Glu Ser Gln Cys Thr
20 25 30

Leu Leu Asn Ala Ser Ile Thr Glu Thr Phe Asn Cys Ser Phe Ser Cys
35 40 45

Gly Pro Asp Cys Trp Lys Leu Ser Gln Tyr Pro Cys Leu Gln Val Tyr 50 55 60

Val Asn Leu Thr Ser Ser Gly Glu Lys Leu Leu Leu Tyr His Thr Glu 65 70 75 80

Glu Thr Ile Lys Ile Asn Gln Lys Cys Ser Tyr Ile Pro Lys Cys Gly
85 90 95

Lys Asn Phe Glu Glu Ser Met Ser Leu Val Asn Val Val Met Glu Asn 100 105 110

Phe Arg Lys Tyr Gln His Phe Ser Cys Tyr Ser Asp Pro Glu Gly Asn 115 120 125

Gln Lys Ser Val Ile Leu Thr Lys Leu Tyr Ser Ser Asn Val Leu Phe 130 135 140

His Ser Leu Phe Trp Pro Thr Cys Met Met Ala Gly Gly Val Ala Ile 145 150 155 160

Val Ala Met Val Lys Leu Thr Gln Tyr Leu Ser Leu Leu Cys Glu Arg 165 170 175

Ile Gln Arg Ile Asn Arg Xaa 180

<210> 134

<211> 147

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (147)

<223> Xaa equals stop translation

<400> 134

Met Trp Lys Leu Trp Arg Ala Glu Glu Gly Ala Ala Ala Leu Gly Gly

1 10 15

Ala Leu Phe Leu Leu Leu Phe Ala Leu Gly Val Arg Gln Leu Leu Lys 20 25 30

Gln Arg Arg Pro Met Gly Phe Pro Pro Gly Pro Pro Gly Leu Pro Phe 35 40 45

Ile Gly Asn Ile Tyr Ser Leu Ala Ala Ser Ser Glu Leu Pro His Val

Tyr Met Arg Lys Gln Ser Gln Val Tyr Gly Glu Val Gln Pro Arg Arg 65 70 75 80

Ala Pro Gly Arg Glu Gly Arg Gln Ala Gly Pro Gly Trp Pro Gly Pro 85 90 95

Ser Trp Leu Asp Leu Trp Pro Pro Leu Gly Arg Leu Val Gly Thr Ser 100 105 110

Pro Cys Ala Gly Cys Pro Leu Arg Asp Thr Arg Phe Pro Gly Leu Glu 115 120 125

Gly Arg Ser Pro Arg Arg Arg Ala Pro Leu Gln Gly Glu Pro Arg Pro
130 135 140

Cys Arg Xaa 145

<210> 135

<211> 122

<212> PRT

<213> Homo sapiens

<400> 135

Met Arg Val Arg Ile Gly Leu Thr Leu Leu Leu Cys Ala Val Leu Leu 1 5 10 15

Ser Leu Ala Ser Ala Ser Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser 20 25 30

Leu Asp Ser Lys Thr Thr Leu Thr Ser Asp Glu Ser Val Lys Asp His 35

Thr Thr Ala Gly Arg Val Val Ala Gly Gln Ile Phe Leu Asp Ser Glu 50 55 60

Glu Ser Glu Leu Glu Ser Ser Ile Gln Glu Glu Glu Asp Ser Leu Lys

70

75

80

Ser Gln Glu Gly Glu Ser Val Thr Glu Asp Ile Ser Phe Leu Glu Ser 85 90 95

Pro Asn Pro Glu Asn Lys Asp Tyr Glu Glu Pro Lys Lys Val Arg Lys

Pro Gly Ser Leu Asp Ile Phe Leu Ala Phe 115 120

<210> 136

<211> 112

<212> PRT

<213> Homo sapiens

<400> 136

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly
1 5 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly 20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys 35 40 45

Cys Met Asp Cys Ser Thr Ser Cys Pro Leu Pro Ala Ala Leu Ala His 50 55 60

Pro Trp Gly Arg Ser Glu Pro Asp Leu Arg Ala Gly Ala Ala Phe Trp 65 70 75 80

Leu Phe Gly Leu Glu Thr Met Pro Gln Arg Glu Lys Phe Thr Thr Pro 85 90 95

Ile Glu Glu Thr Gly Gly Glu Gly Cys Pro Ala Val Ala Leu Ile Gln
100 105 110

<210> 137

<211> 140

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (140)

<223> Xaa equals stop translation

<400> 137

Met Leu Leu Gly Pro Val Pro Ile Leu His Ile Lys Ser Gln Leu Trp

1 10 15

Leu Leu Val Leu Ile Leu Val Val Ser Gly Leu Ser Ala Gly Met Ser 20 25 30

Ile Ile Pro Thr Phe Pro Glu Ile Leu Ser Cys Ala His Glu Asn Gly 35 40 45

Phe Glu Glu Gly Leu Ser Thr Leu Gly Leu Val Ser Gly Leu Phe Ser 50 60

Ala Met Trp Ser Ile Gly Ala Phe Met Gly Pro Thr Leu Gly Gly Phe 65 70 75 80

Leu Tyr Glu Lys Ile Gly Phe Glu Trp Ala Ala Ala Ile Gln Gly Leu 85 90 95

Trp Ala Leu Ile Ser Gly Leu Ala Met Gly Leu Phe Tyr Leu Leu Glu 100 105 110

- Tyr Ser Arg Arg Lys Arg Ser Lys Ser Gln Asn Ile Leu Ser Thr Glu 115 120 125
- Glu Glu Arg Thr Thr Leu Leu Pro Asn Glu Thr Xaa 130 135 140
- <210> 138
- <211> 404
- <212> PRT
- <213> Homo sapiens
- <400> 138
- Met Arg Leu Gln Asp Val Tyr Met Leu Asn Val Lys Gly Leu Ala Arg

 1 5 10 15
- Gly Val Phe Gln Arg Val Thr Gly Ser Ala Ile Thr Asp Leu Tyr Ser 20 25 30
- Pro Lys Arg Leu Phe Ser Leu Thr Gly Asp Asp Cys Phe Gln Val Gly 35 40 45
- Lys Val Ala Tyr Asp Met Gly Asp Tyr Tyr His Ala Ile Pro Trp Leu 50 60
- Glu Glu Ala Val Ser Leu Phe Arg Gly Ser Tyr Gly Glu Trp Lys Thr
 65 70 75 80
- Glu Asp Glu Ala Ser Leu Glu Asp Ala Leu Asp His Leu Ala Phe Ala 85 90 95
- Tyr Phe Arg Ala Gly Asn Val Ser Cys Ala Leu Ser Leu Ser Arg Glu
 100 105 110
- Phe Leu Leu Tyr Ser Pro Asp Asn Lys Arg Met Ala Arg Asn Val Leu 115
- Lys Tyr Glu Arg Leu Leu Ala Glu Ser Pro Asn His Val Val Ala Glu 130 135 140
- Ala Val Ile Gln Arg Pro Asn Ile Pro His Leu Gln Thr Arg Asp Thr 145 150 155 160
- Tyr Glu Gly Leu Cys Gln Thr Leu Gly Ser Gln Pro Thr Leu Tyr Gln
 165 170 175
- Ile Pro Ser Leu Tyr Cys Ser Tyr Glu Thr Asn Ser Asn Ala Tyr Leu 180 185 190
- Leu Leu Gln Pro Ile Arg Lys Glu Val Ile His Leu Glu Pro Tyr Ile 195 200 205
- Ala Leu Tyr His Asp Phe Val Ser Asp Ser Glu Ala Gln Lys Ile Arg 210 215 220
- Glu Leu Ala Glu Pro Trp Leu Gln Arg Ser Val Val Ala Ser Gly Glu 225 235 240

99

Lys Gln Leu Gln Val Glu Tyr Arg Ile Ser Lys Ser Ala Trp Leu Lys 245 250 255

Asp Thr Val Asp Leu Lys Leu Val Thr Leu Asn His Arg Ile Ala Ala 260 265 270

Leu Thr Gly Leu Asp Val Arg Pro Pro Tyr Ala Glu Tyr Leu Gln Val 275 280 285

Val Asn Tyr Gly Ile Gly Gly His Tyr Glu Pro His Phe Asp His Ala 290 295 300

Thr Ser Pro Ser Ser Pro Leu Tyr Arg Met Lys Ser Gly Asn Arg Val 315 320

Ala Thr Phe Met Ile Tyr Leu Ser Ser Val Glu Ala Gly Gly Ala Thr 325 330 335

Ala Phe Ile Tyr Ala Asn Leu Ser Val Pro Val Val Arg Asn Ala Ala 340 345 350

Leu Phe Trp Trp Asn Leu His Arg Ser Gly Glu Gly Asp Ser Asp Thr 355 360 365

Leu His Ala Gly Cys Pro Val Leu Val Gly Asp Lys Trp Val Ala Asn 370 375 380

Lys Trp Ile His Glu Tyr Gly Gln Glu Phe Arg Arg Pro Cys Ser Ser 385 390 395 400

Ser Pro Glu Asp

<210> 139

<211> 96

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (96)

<223> Xaa equals stop translation

<400> 139

Met Lys Ala Pro His Thr Gly Val Leu His Leu Gly Ser Val Trp Val 1 5 10 15

Phe Leu Gly Pro Phe Leu Leu Gly Val Gly Tyr Thr Leu Thr Phe Asn 20 25 30

Pro Leu Ser Gly Cys Met Ser Thr Val Arg Trp Leu Asn Ser Asn Ile 35 40 45

Thr Ala Asn Arg Thr Leu Ser Arg Ser Val Cys His Val Thr Pro Leu 50 60

His Arg Ser Leu Ser Pro His Asp Gly Glu Tyr Leu Arg Gln Met Leu 65 70 75 80

Leu Asn Ser Ser Ser Arg Ala Gly Glu Ala Gly Ser Trp Gly Tyr Xaa 85 90 95

<210> 140 <211> 240 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (240) <223> Xaa equals stop translation <400> 140 Met Gly Ser Cys Ala Arg Leu Leu Leu Trp Gly Cys Thr Val Val Ala Ala Gly Leu Ser Gly Val Ala Gly Val Ser Ser Arg Cys Glu Lys Ala Cys Asn Pro Arg Met Gly Asn Leu Ala Leu Gly Arg Lys Leu Trp 40 Ala Asp Thr Thr Cys Gly Gln Asn Ala Thr Glu Leu Tyr Cys Phe Tyr 55 Ser Glu Asn Thr Asp Leu Thr Cys Arg Gln Pro Lys Cys Asp Lys Cys Asn Ala Ala Tyr Pro His Leu Ala His Leu Pro Ser Ala Met Ala Asp 85 Ser Ser Phe Arg Phe Pro Arg Thr Trp Trp Gln Ser Ala Glu Asp Val 105 His Arg Glu Lys Ile Gln Leu Asp Leu Glu Ala Glu Phe Tyr Phe Thr 120 His Leu Ile Val Met Phe Lys Ser Pro Arg Pro Ala Ala Met Val Leu 135 Asp Arg Ser Gln Asp Phe Gly Lys Thr Trp Lys Pro Tyr Lys Tyr Phe Ala Thr Asn Cys Ser Ala Thr Phe Gly Leu Glu Asp Asp Val Val Lys 165 170 Lys Gly Ala Ile Cys Thr Ser Lys Tyr Ser Ser Pro Phe Pro Cys Thr 185 Gly Arg Lys Val Ile Phe Lys Ala Leu Ser Pro Pro Tyr Asp Thr Glu 195 200 205

Asn Pro Tyr Ser Ala Lys Val Gln Glu Gln Leu Lys Ile Thr Asn Leu

WO 00/06698 PCT/US99/17130

101

210 215 220

Pro Arg Ala Ala Ala Glu Thr Thr Val Leu Ser Leu Ser Glu Lys Xaa 225 230 235 240

<210> 141

<211> 54

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (54)

<223> Xaa equals stop translation

<400> 141

Met Met Ile Ser Gly Leu Lys Leu Leu Val Leu Phe Leu Lys Phe Ala 1 5 10 15

Pro Glu Asn Tyr Cys Leu Ser Thr Glu Thr Leu Gln Met Pro Asn Arg 20 25 30

His Leu Arg Leu Ser Lys Ala Thr Cys Tyr Leu Met Lys Cys Leu Leu $35 \hspace{1cm} 40 \hspace{1cm} 45$

Pro Ser Tyr Phe Glu Xaa 50

<210> 142

<211> 67

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (67)

<223> Xaa equals stop translation

<400> 142

Met Arg Ser Leu Ile Ser Ser His Pro Cys Gln His Leu Leu Leu 1 5 10

Leu Leu Leu Phe Leu Ile Leu Ala Ile Leu Val Asp Val Lys Trp

Tyr Leu Val Leu Phe Ile Cys Ile Ser Leu Met Thr Ser Asp Val Glu 35 40

His Leu Phe Met Cys Leu Leu Ala Ile Arg Ile Ser Ser Trp Arg Asn 50 55

Val Tyr Xaa

65

```
<210> 143
 <211> 108
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (48)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (55)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (58)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (67)
<223> Xaa equals any of the naturally occurring L-amino acids
Met Phe Tyr Lys Leu Thr Leu Ile Leu Cys Glu Leu Ser Val Ala Gly
                                      10
Val Thr Gln Ala Ala Ser Gln Arg Pro Leu Gln Arg Leu Pro Arg His
Ile Cys Ser Gln Arg Asn Pro Pro Gly Arg Cys Leu Leu Lys Ala Xaa
Leu Gln Thr Thr Trp Gly Xaa Pro Asp Xaa Gln Phe Pro Gly Cys Pro
                          55
                                              60
His Pro Xaa Arg Val Thr Leu Asn Ala Arg Gln Met Gly Asn Gly Lys
Glu Lys Lys Ala Ala Asp Leu Lys Leu Lys Phe Pro Gln Lys Arg Phe
Tyr Leu Ser Ala Phe Ser Glu Arg Ile Lys Ala Phe
            100
<210> 144
<211> 84
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (84)
```

<223> Xaa equals stop translation

<400> 144

Met Ala Ser Val Gly Thr Thr Leu Val Ser Pro Leu Leu Cys Leu Leu 1 5 10 15

Ile Pro Thr Arg Val Ser Asp Pro Trp Leu Gln Asn Thr Pro Leu His 20 25 30

Pro Trp Lys Thr lle Thr lle Ile Asp Tyr Tyr Leu Ser Leu Gly Phe 35 40 45

Leu Gly Trp Thr Gly Leu Ser Trp Val Val His Phe Gly Ala Ser Ala 50 60

Val Met Gly Arg Gln Trp Leu Gly Ser Leu Gln Arg Leu Pro Cys Ile 65 70 75 80

Ser Gly Ser Xaa

<210> 145

<211> 166

<212> PRT

<213> Homo sapiens

<400> 145

Met Gly Ser Arg Phe Leu Leu Val Leu Leu Ser Gly Leu Thr Val Leu 1 5 10 15

Leu Ala Leu Pro Gly Ser Glu Ala Lys Asn Ser Gly Ala Ser Cys Pro 20 25 30

Pro Cys Pro Lys Tyr Ala Ser Cys His Asn Ser Thr His Cys Thr Cys 35 40

Glu Asp Gly Phe Arg Ala Arg Ser Gly Arg Thr Tyr Phe His Asp Ser 50 60

Ser Glu Lys Cys Glu Asp Ile Asn Glu Cys Glu Thr Gly Leu Ala Lys 65 70 75 80

Cys Lys Tyr Lys Ala Tyr Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys
85

Ser Cys Leu Val Lys Tyr Thr Leu Phe Asn Phe Leu Ala Gly Ile Ile 100 105 110

Asp Tyr Asp His Pro Asp Cys Tyr Glu Asn Asn Ser Gln Gly Thr Thr 115 120 125

Gln Ser Asn Val Asp Ile Trp Val Ser Gly Val Lys Pro Gly Phe Gly 130 135 140

Lys Gln Leu Val Arg Ile Thr Met Pro Phe Ser Tyr Pro Asn Ile Asn 145 155 160

Met Ser Ser Cys Asp Phe 165 <210> 146 <211> 70 <212> PRT

<213> Homo sapiens

<400> 146

Met Lys Pro Lys His Leu Glu Trp Cys Leu Ala His Ser Trp Cys Val 1 5 10 15

Ile Trp Leu Ser Phe Val Ser Pro Pro Thr Ser His Leu Glu Cys Asp
20 25 30

Gly Phe Pro Gly Ser Leu Leu Pro Pro Cys Glu Glu Gly Arg Cys Phe 35 40 45

Pro Phe Thr Phe His His His Asp Cys His Gly Cys Ser Pro Leu Gln 50 55 60

Ser Ser Pro Gly Gln His 65 70

<210> 147

<211> 412

<212> PRT

<213> Homo sapiens

<400> 147

Met Cys Cys Trp Pro Leu Leu Leu Leu Trp Gly Leu Leu Pro Gly Thr
1 5 10 15

Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu Leu Asp
20 25 30

Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly Ser Gln Ile 35 40 45

Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val Gly Phe Gly Phe 50 60

Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile Val Val Gly Gly Val 65 70 75 80

Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr Phe Thr Asn Ala Asn Arg 85 90 95

Glu Leu Lys Lys Asp Ala Gln Gln Asp Tyr His Leu Glu Tyr Ala Met 100 105 110

Glu Asn Ser Thr His Thr Ile Ile Glu Phe Thr Arg Glu Leu His Thr 115 120 125

Cys Asp Ile Asn Asp Lys Ser Ile Thr Asp Ser Thr Val Arg Val Ile 130 135 140

Trp Ala Tyr His His Glu Asp Ala Gly Glu Ala Gly Pro Lys Tyr His 145 150 155 160

- Asp Ser Asn Arg Gly Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys
 165 170 175
- Thr Ser Val Leu Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln 180 185 190
- Asp Val Pro Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe 195 200 205
- Lys Ile Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro 210 215 220
- Val Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr 225 230 235 240
- Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Leu Glu Ser Gly His Glu 245 250 255
- Cys Tyr His Pro Asn Met Pro Asp Ala Phe Leu Thr Cys Glu Thr Val 260 265 270
- Ile Phe Ala Trp Ala Ile Gly Gly Glu Gly Phe Ser Tyr Pro Pro His 275 280 285
- Val Gly Leu Ser Leu Gly Thr Pro Leu Asp Pro His Tyr Val Leu Leu 290 295 300
- Glu Val His Tyr Asp Asn Pro Thr Tyr Glu Glu Gly Leu Ile Asp Asn 305 310 315 320
- Ser Gly Leu Arg Leu Phe Tyr Thr Met Asp Ile Arg Lys Tyr Asp Ala 325 330 335
- Gly Val Ile Glu Ala Gly Leu Trp Val Ser Leu Phe His Thr Ile Pro $340 \hspace{1cm} 345 \hspace{1cm} 350 \hspace{1cm}$
- Pro Gly Met Pro Glu Phe Gln Ser Glu Gly His Cys Thr Leu Glu Cys 355 360 365
- Leu Glu Glu Leu Trp Lys Pro Lys Ser Gln Val Glu Phe Met Cys Leu 370 380
- Leu Phe Phe Ser Met Leu Thr Trp Leu Ala Glu His Gln Ala Ala Ser 385 390 395 400
- Phe Ser Lys Arg Glu Gly Asn Glu Ile Thr Cys Leu
 405
- <210> 148
- <211> 85
- <212> PRT
- <213> Homo sapiens
- <220>
- <221> SITE
- <222> (85)
- <223> Xaa equals stop translation

<400> 148

Met Asn Val Phe Leu Pro Pro Ala Leu Gly Thr Trp Gly Val Ala Arg

1 5 10 15

Phe Phe Pro His Leu Val Pro Glu Arg Trp Cys Leu Val Phe Cys Cys 20 25 30

Trp Ile Phe Phe Phe Phe Phe Phe Phe Cys Thr Lys Val Ala Thr Arg 35 40 45

Ser Val Leu Gly Asp Gln Ala Gly Leu Gly Val Gly Gly Pro His Leu 50 60

Pro Leu Pro Gly Ser His Ser Val Ser Val Pro Glu Lys Thr Ile Phe 65 70 75 80

Ser Leu Lys Gln Xaa

<210> 149

<211> 154

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (154)

<223> Xaa equals stop translation

<400> 149

Met Gly Arg Leu Pro Leu Leu Arg Arg Val Leu Lys Gly Leu Gln Leu 1 5 10 15

Leu Leu Ser Leu Leu Ala Phe Ile Cys Glu Glu Val Val Ser Gln Cys 20 25 30

Thr Leu Cys Gly Gly Leu Tyr Phe Phe Glu Phe Val Ser Cys Ser Ala 35 40 45

Phe Leu Leu Ser Leu Leu Ile Leu Ile Val Tyr Cys Thr Pro Phe Tyr 50 60

Glu Arg Val Asp Thr Thr Lys Val Lys Ser Ser Asp Phe Tyr Ile Thr 65 70 75 80

Leu Gly Thr Gly Cys Val Phe Leu Leu Ala Ser Ile Ile Phe Val Ser 85 90 95

Thr His Asp Arg Thr Ser Ala Glu Ile Ala Ala Ile Val Phe Gly Phe 100 105 110

Ile Ala Ser Phe Met Phe Leu Leu Asp Phe Ile Thr Met Leu Tyr Glu 115 120 125

Lys Arg Gln Glu Ser Gln Leu Arg Lys Pro Glu Asn Thr Thr Arg Ala 130 135 140

Glu Ala Leu Thr Glu Pro Leu Asn Ala Xaa .

107

145 150 <210> 150 <211> 130 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (130) <223> Xaa equals stop translation <400> 150 Met Arg Gly His Leu Ala Gly Phe Pro Ala Leu Ser Gly Leu Ala Ser 10 Val Cys Leu Trp Ala Thr Phe Ser Ala Gln Leu Pro Gly Pro Val Ala Ala Thr Ser Trp Thr Pro Ala Pro Leu Gly Cys Ser Ala Ala Arg Ser Gly Pro Glu Lys Arg Leu Gly Thr Ala Ala Pro Gly Ser Ala Ala Ser Leu Ala Gln Ala Gly Pro Gly Ala Pro Cys Arg Val Leu Pro Val Asp Pro Ala Pro Ala Ala Leu Asn Val Arg Glu Pro Gly Trp Leu Gly Gly 90 Leu Phe Asp Gly Ala Leu Leu Gln Val Leu Leu Asn Phe Leu Arg Lys 100 105 Ser Thr Asp Val Leu Met Asp Thr Arg Glu Ala Glu Ser Leu Glu Val 120 Glu Xaa 130 <210> 151 <211> 62 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (62) <223> Xaa equals stop translation <400> 151 Met Leu Phe Trp Ala Tyr Pro Ile Cys Val Phe Ile Asp Ser Leu Ser Cys Gln Pro Cys Leu Trp Ser Thr Gly Ala Thr Ser His Phe Asn Ser

Pro Thr Thr Ser Pro Leu Phe Thr Leu Phe Met Pro Cys Ala Leu Ala 35 40 45

Pro Asn Pro Phe Thr Gln Leu Gly Lys Leu Asp Asp Arg Xaa 50 55 60

<210> 152

<211> 225

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (225)

<223> Xaa equals stop translation

<400> 152

Met Gly Ile Phe Pro Gly Ile Ile Leu Ile Phe Leu Arg Val Lys Phe 1 5 10 15

Ala Thr Ala Ala Val Ile Val Ser Gly His Gln Lys Ser Thr Thr Val 20 25 30

Ser His Glu Met Ser Gly Leu Asn Trp Lys Pro Phe Val Tyr Gly Gly 35 40 45

Leu Ala Ser Ile Val Ala Glu Phe Gly Thr Phe Pro Val Asp Leu Thr . 50 55 60

Lys Thr Arg Leu Gln Val Gln Gly Gln Ser Ile Asp Ala Arg Phe Lys 65 70 75 80

Glu Ile Lys Tyr Arg Gly Met Phe His Ala Leu Phe Arg Ile Cys Lys
85 90 95

Glu Glu Gly Val Leu Ala Leu Tyr Ser Gly Ile Ala Pro Ala Leu Leu 100 105 110

Arg Gln Ala Ser Tyr Gly Thr Ile Lys Ile Gly Ile Tyr Gln Ser Leu 115 120 125

Lys Arg Leu Phe Val Glu Arg Leu Glu Asp Glu Thr Leu Leu Ile Asn 130 135 140

Pro Thr Asp Val Leu Lys Ile Arg Met Gln Ala Gln Gly Ser Leu Phe 165 170 175

Gin Gly Ser Met Ile Gly Ser Phe Ile Asp Ile Tyr Gln Gln Glu Gly 180 185 190

Thr Arg Gly Leu Trp Arg Val Ser Thr Leu Phe Leu Leu Leu Ser Tyr 195 200 205

Thr Leu Ser Ser Tyr Asn Leu Gln Arg Ile Phe Phe Tyr Ile Lys Thr 210 215 220

Ser Tyr Val Xaa

```
Xaa
225
<210> 153
<211> 69
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (69)
<223> Xaa equals stop translation
<400> 153
Met Leu Met Leu Leu Thr Leu Leu Val Leu Gly Met Val Trp Val Ala
Ser Ala Ile Val Asp Lys Asn Lys Ala Asn Arg Glu Ser Leu Tyr Asp
                                  25
Phe Trp Glu Tyr Tyr Leu Pro Tyr Leu Tyr Ser Cys Ile Ser Phe Leu
Gly Val Leu Leu Leu Ala Ala Gly Arg Pro Gly Gly Ala Ala Val
Leu Leu Ser Leu Xaa
 65
<210> 154
<211> 84
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (84)
<223> Xaa equals stop translation
<400> 154
Met Tyr Gly Val Cys Leu Cys Val Ile Val Cys Val Ser Gly Val Ser
Leu Cys Leu Tyr Val Trp Gly Val Ser Val Cys Asp Cys Val Ser Val
                                 25
Phe Met Cys Val Cys Leu Cys Val Ile Phe Cys Val Tyr Gly Lys Pro
                             40
Arg Thr Glu His Tyr His Ser Pro His Leu Ala Lys Gln Lys Ala Phe
Arg Glu Met Cys Gly Arg His Asp Val Ser Ala Ala Gly Ile Phe Gln
65
```

<210> 157 <211> 162

```
<210> 155
 <211> 61
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (61)
 <223> Xaa equals stop translation
 <400> 155
 Met His Val Leu Leu Phe Ser Phe Leu Ile Pro Phe Leu Leu Leu Ser
                                      10
 Pro Val Gly Val Thr Cys Asn Ser His Met Leu Glu Arg Gln Val Ser
              20
 Trp Leu Lys Lys Arg Ser Thr Gln Ala Ser Gln Gln Phe Asn Lys Phe
                              40
 Leu Arg Gly Ile Ser Asn Val Gly Arg Ile Val Ile Xaa
 <210> 156
 <211> 84
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (84)
<223> Xaa equals stop translation
<400> 156
Met Cys Leu Leu Val Glu Tyr Ser Leu Met Ile Leu Thr Ile Ile Pro
Ser Leu Leu Ser Phe Val Leu Cys Leu Lys Gly Ile Lys His Gly Asn
Tyr Ile Phe Gln Thr Pro Leu Pro Glu Gly Tyr Gly Trp Ile Ser Ala
Met Ser Gly Leu Cys Ile Lys Phe Gly Arg Arg Lys Arg Lys Thr
Trp Leu Leu Gln Val Gly Thr Leu Ala Thr Ile Asp Thr Glu Phe Ala
                     70
Arg Ser Cys Xaa
```

<220>
<221> SITE
<222> (107)

<220>

```
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (162)
<223> Xaa equals stop translation
<400> 157
Met Ala Leu Ser Leu Thr Leu Cys Phe Val Met Phe Trp Thr Pro Asn
Val Ser Glu Lys Ile Leu Ile Asp Ile Ile Gly Val Asp Phe Ala Phe
                                  25 .
Ala Glu Leu Cys Val Val Pro Leu Arg Ile Phe Ser Phe Phe Pro Val
                              40
Pro Val Thr Val Arg Ala His Leu Thr Gly Trp Leu Met Thr Leu Lys
Lys Thr Phe Val Leu Ala Pro Ser Ser Val Leu Arg Ile Ile Val Leu
Ile Ala Ser Leu Val Val Leu Pro Tyr Leu Gly Val His Gly Ala Thr
Leu Gly Val Gly Ser Leu Leu Ala Gly Phe Val Gly Glu Ser Thr Met
                                 105
Val Ala Ile Ala Ala Cys Tyr Val Tyr Arg Lys Gln Lys Lys Met
        115
                            120
Glu Asn Glu Ser Ala Thr Glu Gly Glu Asp Ser Ala Met Thr Asp Met
                        135
                                            140
Pro Pro Thr Glu Glu Val Thr Asp Ile Val Glu Met Arg Glu Glu Asn
Glu Xaa
<210> 158
<211> 146
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (96)
<223> Xaa equals any of the naturally occurring L-amino acids
```

<223> Xaa equals any of the naturally occurring L-amino acids

```
<221> SITE
  <222> (111)
  <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
  <221> SITE
  <222> (115)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (122)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (132)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <400> 158
 Met Glu Pro Gln Leu Gly Pro Glu Ala Ala Ala Leu Arg Pro Gly Trp
 Leu Ala Leu Leu Trp Val Ser Ala Leu Ser Cys Ser Phe Ser Leu
             20
 Pro Ala Ser Ser Leu Ser Ser Leu Val Pro Gln Val Arg Thr Ser Tyr
                              40
Asn Phe Gly Arg Thr Phe Leu Gly Leu Asp Lys Cys Asn Ala Cys Ile
Gly Thr Ser Ile Cys Lys Lys Phe Phe Lys Glu Glu Ile Arg Ser Asp
 65
Asn Trp Leu Ala Ser His Leu Gly Thr Ala Ser Arg Phe Pro Leu Xaa
Ser Tyr Pro Cys Lys Leu Leu Gln Met Ile Xaa Lys Ile Trp Xaa Pro
            100
                                 105
                                                     110
Cys Gly Xaa Leu Leu Thr Gly Gln Gln Xaa Ser Asn Glu Ile Ser Lys
                            120
Gln Glu Ile Xaa Cys Leu Leu His Pro Pro Pro Lys Asn Leu His Ile
    130
                        135
Asp Val
145
<210> 159
<211> 143
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (143)
```

<223> Xaa equals stop translation

<400> 159

Met Trp Trp Ala Val Met Gly Gly Val Ile Gly Ser Trp Leu Ser Pro 1 5 10

Leu Ser Ile Ala Glu Cys Cys His Asp Leu Trp Thr Ser Gln Ser Cys 20 25 30

Glu. His Ala Gly Ala Leu Cys Gly Asp Leu Leu Cys Ala Cys Arg Lys 35 40

Val Gly Val Trp Cys Ala Leu Gln Gln His Trp Trp Asn Arg Cys Val 50 60

Cys Pro His Ala Val Ile Arg Val His Cys Thr Gly Ala Ser Tyr Thr
65 70 75 80

Leu Gln Lys Ile Cys Ser Cys Asn Pro Lys Phe Met Gly Arg His Pro 85 90 95

His Arg Trp Gln Gln Ile Arg Lys Cys Ser Gln Pro Val Leu Arg Gly
100 105 110

Ser Arg Ala Ala Phe Ile Trp Val Arg Leu Ala Ala Leu Asn Phe Ile 115 120 125

Ser Ser Phe Arg Cys Ile Ser Leu Ile Ser Tyr Ser Ala Phe Xaa 130 140

<210> 160

<211> 51

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (51)

<223> Xaa equals stop translation

<400> 160

Met Lys Val Ser Asp Phe Asn Phe Leu Ile Phe Leu Ile Phe Ala Leu

1 5 10 15

Phe Leu Thr Leu Glu Ala Phe Leu Lys Phe Thr Lys Arg Val Leu Ala 20 25 30

Val Val Gly Asn Leu Pro Glu Pro Pro Ile Ile Lys Thr Ile Gly Phe

Leu Tyr Xaa 50

<210> 161

<211> 65

<212> PRT

<213> Homo sapiens

```
<220>
 <221> SITE
 <222> (65)
 <223> Xaa equals stop translation
 <400> 161
 Met Val Trp Ser Ala Ala Pro Ala Pro Cys Cys Leu Leu Gly Val Leu
                                       10
 Gly Leu Val Gln Val Leu Gly Ala Gln Ala Val Gly Pro Trp Thr Ala
 Ser Ala Cys Leu Gly Ala Ala Gln Ala Gln Pro Cys Arg Pro Cys Lys
                               40
 Glu Ser Ser Leu Arg Leu Phe Ser Ala Ser Ala Pro Ser Met Thr His
                          55
 Xaa
  65
 <210> 162
 <211> 59
 <212> PRT
 <213> Homo sapiens
 <220>
<221> SITE
 <222> (59)
 <223> Xaa equals stop translation
<400> 162
Met Glu Lys Tyr Cys Leu Gly Asn Asn Met Leu Ser Arg Phe Cys Leu
                  5
Phe Leu Ile Met Leu Leu His Ile Leu Leu Phe Leu Val Ile Phe Ile
Gln Arg His Thr Val Val Ser Leu Ser Lys His His Pro Phe Val Pro
                              40
Thr Asn Gly Ser Lys Ser Tyr Ser Ser Phe Xaa
<210> 163
<211> 374
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (84)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
```

<222>	(1	1	2)
-------	---	---	---	---	---

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 163

Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val Leu Leu $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln Pro Val 20 25 30

Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr Phe His $35 \hspace{1cm} 40 \hspace{1cm} 45$

Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala Cys Arg 50 60

Arg Asp Gly Gly Gln Leu Val Ser Ile Glu Ser Glu Asp Glu Gln Lys 65 70 75 80

Leu Ile Glu Xaa Phe Ile Glu Asn Leu Leu Pro Ser Asp Gly Asp Phe
85 90 95

Trp Ile Gly Leu Arg Arg Glu Glu Lys Gln Ser Asn Ser Thr Xaa 100 105 110

Cys Gln Asp Leu Tyr Ala Trp Thr Asp Gly Ser Ile Ser Gln Phe Arg 115 120 125

Asn Trp Tyr Val Asp Glu Pro Ser Cys Gly Ser Glu Val Cys Val Val 130 140

Met Tyr His Gln Pro Ser Ala Pro Ala Gly Ile Gly Gly Pro Tyr Met 145 150 155 160

Phe Gln Trp Asn Asp Asp Arg Cys Asn Met Lys Asn Asn Phe Ile Cys 165 170 175

Lys Tyr Ser Asp Glu Lys Pro Ala Val Pro Ser Arg Glu Ala Glu Gly 180 185 190

Glu Glu Thr Glu Leu Thr Thr Pro Val Leu Pro Glu Glu Thr Gln Glu
195 200 205

Glu Asp Ala Lys Lys Thr Phe Lys Glu Ser Arg Glu Ala Ala Leu Asn 210 215 220

Leu Ala Tyr Ile Leu Ile Pro Ser Ile Pro Leu Leu Leu Leu Leu Val 225 235 240

Val Thr Thr Val Val Cys Trp Val Trp Ile Cys Arg Lys Arg Lys Arg 255 255

Glu Gln Pro Asp Pro Ser Thr Lys Lys Gln His Thr Ile Trp Pro Ser 260 265 270

Pro His Gln Gly Asn Ser Pro Asp Leu Glu Val Tyr Asn Val Ile Arg 275 280 285

Lys Gln Ser Glu Ala Asp Leu Ala Glu Thr Arg Pro Asp Leu Lys Asn

290 295 300 Ile Ser Phe Arg Val Cys Ser Gly Glu Ala Thr Pro Asp Asp Met Ser 310 315 Cys Asp Tyr Asp Asn Met Ala Val Asn Pro Ser Glu Ser Gly Phe Val 330 Thr Leu Val Ser Val Glu Ser Gly Phe Val Thr Asn Asp Ile Tyr Glu 340 345 Phe Ser Pro Asp Gln Met Gly Arg Ser Lys Glu Ser Gly Trp Val Glu 360 365 Asn Glu Ile Tyr Gly Tyr 370 <210> 164 <211> 64 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (64) <223> Xaa equals stop translation <400> 164 Met His Pro Gln Leu Ile Pro Ser Val Ile Ala Val Val Phe Ile Leu Leu Leu Gly Val Cys Phe Ile Ala Ser Cys Leu Val Thr His His Asn 20 Phe Ser Arg Cys Lys Arg Gly Thr Gly Val His Lys Leu Glu His His Ala Lys Leu Lys Cys Ile Lys Glu Lys Ser Glu Leu Lys Ser Cys Xaa <210> 165 <211> 743 <212> PRT <213> Homo sapiens <220> <221> SITE

<222> (743)
<223> Xaa equals stop translation
<400> 165
Met Ala Val Arg Glu Leu Cys Phe Pro Arg Gln Arg Gln Val Leu Phe
1 5 10 15

- Leu Phe Leu Phe Trp Gly Val Ser Leu Ala Gly Ser Gly Phe Gly Arg 20 25 30
- Tyr Ser Val Thr Glu Glu Thr Glu Lys Gly Ser Phe Val Val Asn Leu 35 40
- Ala Lys Asp Leu Gly Leu Ala Glu Gly Glu Leu Ala Arg Gly Thr 50 60
- Arg Val Val Ser Asp Asp Asn Lys Gln Tyr Leu Leu Asp Ser His 65 70 75 80
- Thr Gly Asn Leu Leu Thr Asn Glu Lys Leu Asp Arg Glu Lys Leu Cys
 85
 90
 95
- Gly Pro Lys Glu Pro Cys Met Leu Tyr Phe Gln Ile Leu Met Asp Asp 100 105 110
- Pro Phe Gln Ile Tyr Arg Ala Glu Leu Arg Val Arg Asp Ile Asn Asp 115 120 125
- His Ala Pro Val Phe Gln Asp Lys Glu Thr Val Leu Lys Ile Ser Glu 130 135 140
- Asn Thr Ala Glu Gly Thr Ala Phe Arg Leu Glu Arg Ala Gln Asp Pro 145 150 155 160
- Asp Gly Gly Leu Asn Gly Ile Gln Asn Tyr Thr Ile Ser Pro Asn Ser 165 170 175
- Phe Phe His Ile Asn Ile Ser Gly Gly Asp Glu Gly Met Ile Tyr Pro 180 185 190
- Glu Leu Val Leu Asp Lys Ala Leu Asp Arg Glu Glu Gln Gly Glu Leu 195 200 205
- Ser Leu Thr Leu Thr Ala Leu Asp Gly Gly Ser Pro Ser Arg Ser Gly 210 215 220
- Thr Ser Thr Val Arg Ile Val Val Leu Asp Val Asn Asp Asn Ala Pro 225 230 235 240
- Gln Phe Ala Gln Ala Leu Tyr Glu Thr Gln Ala Pro Glu Asn Ser Pro 245 250 255
- Ile Gly Phe Leu Ile Val Lys Val Trp Ala Glu Asp Val Asp Ser Gly 260 265 270
- Val Asn Ala Glu Val Ser Tyr Ser Phe Phe Asp Ala Ser Glu Asn Ile 275 280 285
- Arg Thr Thr Phe Gln Ile Asn Pro Phe Ser Gly Glu Ile Phe Leu Arg 290 295 300
- Glu Leu Leu Asp Tyr Glu Leu Val Asn Ser Tyr Lys Ile Asn Ile Gln 305 310 315 320
- Ala Met Asp Gly Gly Leu Ser Ala Arg Cys Arg Val Leu Val Glu 325 330 335

- Val Leu Asp Thr Asn Asp Asn Pro Pro Glu Leu Ile Val Ser Ser Phe 340 345 350
- Ser Asn Ser Val Ala Glu Asn Ser Pro Glu Thr Pro Leu Ala Val Phe 355 360 365
- Lys Ile Asn Asp Arg Asp Ser Gly Glu Asn Gly Lys Met Val Cys Tyr
 375 380
- Ile Gln Glu Asn Leu Pro Phe Leu Leu Lys Pro Ser Val Glu Asn Phe 385 390 395
- Tyr Ile Leu Ile Thr Glu Gly Ala Leu Asp Arg Glu Ile Arg Ala Glu
 405 410 415
- Tyr Asn Ile Thr Ile Thr Val Thr Asp Leu Gly Thr Pro Arg Leu Lys
 420 425 430
- Thr Glu His Asn Ile Thr Val Leu Val Ser Asp Val Asn Asn Asn Ala 435 440 445
- Pro Ala Phe Thr Gln Thr Ser Tyr Thr Leu Phe Val Arg Glu Asn Asn 450 455 460
- Ser Pro Ala Leu His Ile Gly Ser Val Ser Ala Thr Asp Arg Asp Ser 465 470 475 480
- Gly Thr Asn Ala Gln Val Thr Tyr Ser Leu Leu Pro Pro Gln Asp Pro 485 490 495
- His Leu Pro Leu Ala Ser Leu Val Ser Ile Asn Ala Asp Asn Gly His 500 505 510
- Leu Phe Ala Leu Arg Ser Leu Asp Tyr Glu Ala Leu Gln Ala Phe Glu 515 525
- Phe Arg Val Gly Ala Thr Asp Arg Gly Ser Pro Ala Leu Asn Ser Glu 530 540
- Ala Leu Gly Ala Arg Ala Gly Ala Gly Arg Gln Arg Gln Leu Ala Leu 545 550 555 560
- Arg Ala Val Pro Ala Ala Glu Arg Leu Arg Ala Leu His Arg Ala Gly 565 570 575
- Ala Pro Gly Gly Arg Ala Gly Leu Pro Gly Asp Gln Gly Gly Gly Gly 580 590
- Gly Arg Arg Leu Gly Pro Glu Arg Leu Ala Val Val Pro Ala Ala Gln 595 600 605
- Gly His Gly Ala Arg Ala Val Arg Cys Val Gly Ala Gln Trp Gly Gly 610 615 620
- Ala His Arg Gln Ala Ala Glu Arg Ala Arg Arg Ser Gln Ala Gln Ala 625 630 635 640
- Gly Gly Ala Cys Gln Gly Gln Trp Arg Ala Ser Ser Leu Gly His Arg

645 650 655

His Ala Ala Arg Ala Pro Gly Gly Arg Leu Leu Pro Ala Leu Pro Ala 660 665 670

Ser Pro Gly Gly Gly Pro Gly Pro Gly Arg Leu Ala His Arg 675 680 685

Leu Pro Gly Gly Val Gly Leu Gly Val Phe Ala Leu Pro Pro Leu 690 695 700

Gly Ala Pro Val Arg Gly Gly Ala Ala Val Gln Glu Glu Gln Gly Gly 705 710 715 720

Leu Gly Gly Ser Leu Leu Gly Ala Arg Gly Ser Phe Ser Arg Ala Ser 725 730 735

Gly Gly Arg Glu Gly Arg Xaa 740

<210> 166

<211> 214

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (214)

<223> Xaa equals stop translation

<400> 166

Met Asn Arg Met Glu Leu Leu Lys Leu Leu Leu Thr Cys Phe Ser Glu

1 10 15

Ala Met Tyr Leu Pro Pro Ala Pro Glu Ser Gly Ser Thr Asn Pro Trp
20 25 30

Val Gln Phe Phe Cys Ser Thr Glu Asn Arg His Ala Leu Pro Leu Phe 35 40 45

Thr Ser Leu Leu Asn Thr Val Cys Ala Tyr Asp Pro Val Gly Tyr Gly 50 60

Ile Pro Tyr Asn His Leu Leu Phe Ser Asp Tyr Arg Glu Pro Leu Val 65 70 75 80

Glu Glu Ala Ala Gln Val Leu Ile Val Thr Leu Asp His Asp Ser Ala 85 90

Ser Ser Ala Ser Pro Thr Val Asp Gly Thr Thr Thr Gly Thr Ala Met 100 105 110

Asp Asp Ala Asp Pro Pro Gly Pro Glu Asn Leu Phe Val Asn Tyr Leu 115 120 125

Ser Arg Ile His Arg Glu Glu Asp Phe Gln Phe Ile Leu Lys Gly Ile 130 140 Ala Arg Leu Leu Ser Asn Pro Leu Leu Gln Thr Tyr Leu Pro Asn Ser 155 160

Thr Lys Lys Asp Pro Val Pro Pro Gly Ala Ala Ser Ser Leu Leu Glu 165 170 175

Ala Leu Arg Leu Gln Gln Glu Ile Pro Leu Leu Arg Ala Glu Glu Gln 180 185 190

Arg Arg Pro Arg His Pro Cys Pro His Pro Leu Leu Pro Gln Arg Cys
195 200 205

Pro Gly Arg Ser Val Xaa 210

<210> 167

<211> 213

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (213)

<223> Xaa equals stop translation

<400> 167

Met Pro Ser Leu Arg Phe Leu Ala Leu Ala Leu Leu Leu Ala Ile Leu 1 5 10 15

Pro Ala Leu Pro Asn Ala His Ala Ala Pro Gly Ile Gly Gly Leu Ile
20 25 30

Gly Gly Gly Ser Gln Ala Ser Ala Lys Glu Glu Pro Gln Ser Asn Ala 35 40 45

Gln Pro Ser Ala Asp Glu Arg Lys Gln Arg Leu Leu Ser Gln Ala Glu 50 60

Glu Thr Arg Gln Arg Leu Thr Asp Leu Lys Ala Glu Leu Ala Gly Ala 65 70 75 80

Pro Lys Glu Ile Ser Glu Ala Gln Arg Thr Leu Ser Lys Leu Val Ser 85 90 95

Glu Asp Asn Ser Asp Leu Pro Glu Arg Leu Ser Lys Leu Ser Val Pro
100 105 110

Val Leu Glu Gln Arg Leu Ala Ala Arg Val Asp Glu Leu Ala Leu Trp 115 120 125

Gln Gln Ala Leu Ser Ala Ala Asn Ser Met Leu Ile Ser Ala Gln Thr 130 140

Arg Pro Glu Arg Ala Gln Ala Asp Ile Ser Lys Asn Gln Leu Arg Ile 145 150 155 160

Asp Glu Ile Asn Gly Leu Leu Lys Ser Gly Arg Glu Asn Asn Lys Pro 165 170 175

```
Leu Thr Asp Glu Arg Arg Ala Leu Leu Glu Ser Thr Ser Arg Ala Ala
180 185 190
```

Ala Gly Pro Ser Ile Phe His Pro Gly Gly Val Pro Gly Lys Cys Thr 195 200 205

Gln Phe Ala Leu Xaa · 210

<210> 168

<211> 75

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (75)

<223> Xaa equals stop translation

<400> 168

Met Phe Thr Ser Phe Gly Leu Ala Ser Pro Arg Ile Leu Phe Cys Phe 1 5 10 15

Cys Phe Phe Asp Leu Gly Phe Ile Phe Phe Cys Val Leu Tyr Tyr Ile 20 25 30

Val Lys Gly Ile Leu Ala Glu Thr Leu Val Phe Gly Ala Arg Gly Glu 35 40 45

Gln Glu Cys Trp Ala Val Tyr Phe Arg Trp Arg Thr His Leu Gln Thr 50 55 60

Phe Gly Leu Phe Ser Phe Asn Cys Ser Val Xaa 65 70 75

<210> 169

<211> 48

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (48)

<223> Xaa equals stop translation

<400> 169

Met Phe Leu Cys Leu Phe Phe Phe Phe Phe Asn Ala Thr Gln Gly Asn 1 5 10 15

Ile Phe Ile Ser Phe Leu Ser Gly Leu Pro Gln Cys Ile Phe Ile Ser 20 25 30

Phe Glu Thr Lys Arg Phe Trp Lys Leu Phe Phe Cys Ser Phe Lys Xaa 35 40 45

<210> 172 <211> 48 <212> PRT

<213> Homo sapiens

```
<210> 170
  <211> 88
  <212> PRT
  <213> Homo sapiens
 <220>
 <221> SITE
 <222> (88)
 <223> Xaa equals stop translation
 <400> 170
 Met Gly Leu His Leu Arg Pro Tyr Arg Val Gly Leu Leu Pro Asp Gly
                                      10
 Leu Leu Phe Leu Leu Leu Leu Met Leu Leu Ala Asp Pro Ala Leu
              20
 Pro Ala Gly Arg His Pro Pro Val Val Leu Val Pro Gly Asp Leu Gly
                              40
 Asn Gln Leu Glu Ala Lys Leu Asp Lys Pro Thr Val Val His Tyr Leu
      50
                         55
Cys Ser Lys Lys Thr Glu Ser Tyr Phe Thr Ile Trp Leu Asn Leu Glu
                     70
Leu Leu Pro Val His His Xaa
<210> 171
<211> 42
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (42)
<223> Xaa equals stop translation
<400> 171
Met Ala Cys Glu Thr His Gly Val Leu Val Pro Ala His Leu Ser Gly
Leu Ile Thr Cys Leu Leu Ala Phe Trp Val Pro Ala Ser Cys Ile Gln
             20
Arg Cys Ser Gly Ser Pro Leu Pro Leu Xaa
        35
```

```
123
<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation
Met Gln Cys Phe Leu Phe Ser Ile Phe Leu Ile Thr Gly Leu Ala Glu
                                      10
Glu Phe Cys Glu Gln Leu Ser Ile Ser Leu Ala Glu Glu Glu Ile Gln
             20
Leu Ser Ser Thr Val Glu His Phe Cys Met Thr Ala Phe Ser Trp Xaa
                             40
<210> 173
<211> 233
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (233)
<223> Xaa equals stop translation
<400> 173
Met Ala Ala Leu Ala Ala Ala Lys Lys Val Trp Ser Ala Arg Arg
Leu Leu Val Leu Leu Phe Thr Pro Leu Ala Leu Leu Pro Val Val Phe
Ala Leu Pro Pro Lys Glu Gly Arg Cys Leu Phe Val Ile Leu Leu Met
```

Ala Val Tyr Trp Cys Thr Glu Ala Leu Pro Leu Ser Val Thr Ala Leu

55 Leu Pro Ile Val Leu Phe Pro Phe Met Gly Ile Leu Pro Ser Asn Lys

Val Cys Pro Gln Tyr Phe Leu Asp Thr Asn Phe Leu Phe Leu Ser Gly

Leu Ile Met Ala Ser Ala Ile Glu Glu Trp Asn Leu His Arg Arg Ile 105

Ala Leu Lys Ile Leu Met Leu Val Gly Val Gln Pro Ala Arg Leu Ile 120

Leu Gly Met Met Val Thr Thr Ser Phe Leu Ser Met Trp Leu Ser Asn 130

Thr Ala Ser Thr Ala Met Met Leu Pro Ile Ala Asn Ala Ile Leu Lys 150 155

Ser Leu Phe Gly Gln Lys Glu Val Arg Lys Asp Pro Ser Gln Glu Ser 165 170 175

Glu Glu Asn Thr Gly Ile Glu Pro Asn Thr Phe Leu Ser Glu Glu Arg

Leu Lys Leu Gln Ala Pro Leu Val Ile Arg Leu Gly Gln Ile Thr Glu
195 200 205

Ser Gly Gln Trp Asn Met Ser Gly Asn Asp Val Cys Asn Phe Arg Val 210 215 220

<210> 174

<211> 45

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (45)

<223> Xaa equals stop translation

<400> 174

Met Gly Thr Ile Phe Gly Tyr Leu His Cys Val Lys Cys Tyr Val Leu

1 5 10

Tyr Phe Ile Phe Ile Leu Ile Thr Ala Val Tyr His Ser Phe Tyr Tyr 20 25 30

Pro His Tyr Arg Gly Lys Ala Leu Ile Ser Gly Thr Xaa 35 40 45

<210> 175

<211> 85

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (77)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (85)

<223> Xaa equals stop translation

<400> 175

Met Val Trp Phe Leu Phe Leu Val Phe Ile Phe Leu Lys Val Lys Gly

1 10 15

Asp Phe Phe Pro Pro Phe Leu Ile Cys Asn Leu Phe Cys Ile Trp Met

Ile Thr Gly Val Ser His Arg Leu Gln Pro Gln Ile Leu Phe Ser Arg 35 40 45

His Lys His Asn Gln Glu Ile Ile Leu Gln Met Val Ser Phe Ser Cys 50 55 60

Cys Val Phe Phe Pro Met Ile Arg Glu Val Lys Ser Xaa Leu Gly Cys 65 70 75 80

Ile Lys Met Ser Xaa

<210> 176

<211> 66

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (66)

<223> Xaa equals stop translation

<400> 176

Met Trp Val Leu Leu Ser Cys Pro Leu Pro Pro Leu Cys Leu Pro Ala 1 5 10

Ser Ala Val Pro Gly Gln Cys Leu Gly Gly Gln Trp Ser Gly His Gln 20 25 30

Leu Arg Leu Arg Gly Arg Gly Trp His Cys Arg Cys His Cys Arg Ala 35 40 45

Trp Ala Ala Asp Met Gly Arg Gly Leu His Ser Cys Gln Leu Leu Ser 50 60

Arg Xaa 65

<210> 177

<211> 55

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (55)

<223> Xaa equals stop translation

<400> 177

Met Leu Leu Cys Ile Leu Leu Ile Phe Cys Val Val Gly Leu Ser

Val Val Gly Arg Arg Val Leu Lys Ser Thr Thr Ile Ile Val Tyr Leu 20 25 30

Ser Ile Thr Pro Phe Ser Ser Phe Ser Ser Ile Ser His Ile Phe Gln

```
35 40 45
```

Leu Leu Ile Gly Ala His Xaa 50 55

<210> 178

<211> 83

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (4)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (83)

<223> Xaa equals stop translation

<400> 178

Met Cys Val Xaa Leu Ser Phe Cys Pro Phe Leu Ser Ser Ala Leu Pro
1 5 10

Ala Ser His Thr Gln Phe Tyr Met Pro Arg Gly Ala Lys Phe Gly Thr 20 25 30

Phe Thr Leu Gln Ala Ser Val Ser Pro Leu Glu Glu Lys Thr His Ser 35 40 45

Phe Thr His Pro Gly Ile Gly Gly Lys Leu Leu Gly His Gln Asp Pro 50 55 60

Gly Ala Pro Gly Pro Ser Trp Asn Ile Arg Ser Thr Trp Ser Thr Arg 65 70 75 80

Ser Leu Xaa

<210> 179

<211> 330

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (38)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (247)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 179

Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val 1 5 10 15

Ţ.,...

- Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly 20 25 30
- Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val 35 40
- Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Lys His Leu Ala 50 60
- Arg Leu Gly Met His Val Ile Ile Ala Gly Asn Asn Asp Ser Lys Ala 65 70 75 80
- Lys Gln Val Val Ser Lys Ile Lys Glu Glu Thr Leu Asn Asp Lys Val
- Glu Phe Leu Tyr Cys Asp Leu Ala Ser Met Thr Ser Ile Arg Gln Phe 100 105 110
- Val Gln Lys Phe Lys Met Lys Lys Ile Pro Leu His Val Leu Ile Asn 115 120 125
- Asn Ala Gly Val Met Met Val Pro Gln Arg Lys Thr Arg Asp Gly Phe 130 140
- Glu Glu His Phe Gly Leu Asn Tyr Leu Gly His Phe Leu Leu Thr Asn 145 150 155 160
- Leu Leu Leu Asp Thr Leu Lys Glu Ser Gly Ser Pro Gly His Ser Ala 165 170 175
- Arg Val Val Thr Val Ser Ser Ala Thr His Tyr Val Ala Glu Leu Asn 180 185 190
- Met Asp Asp Leu Gln Ser Ser Ala Cys Tyr Ser Pro His Ala Ala Tyr 195 200 205
- Ala Gln Ser Lys Leu Ala Leu Val Leu Phe Thr Tyr His Leu Gln Arg 210 215 220
- Leu Leu Ala Ala Glu Gly Ser His Val Thr Ala Asn Val Val Asp Pro 225 230 240
- Gly Val Val Asn Thr Asp Xaa Tyr Lys His Val Phe Trp Ala Thr Arg
 245 250 255
- Leu Ala Lys Lys Leu Leu Gly Trp Leu Leu Phe Lys Thr Pro Asp Glu 260 265 270
- Gly Ala Trp Thr Ser Ile Tyr Ala Ala Val Thr Pro Glu Leu Glu Gly 275 280 285
- Val Gly Gly Arg Tyr Leu Tyr Asn Glu Lys Glu Thr Lys Ser Leu His 290 295 300
- Val Thr Tyr Asn Gln Lys Leu Gln Gln Gln Leu Trp Ser Lys Ser Cys 305 310 315 320
- Glu Met Thr Gly Val Leu Asp Val Thr Leu

<221> SITE <222> (55)

<400> 182

<223> Xaa equals stop translation

325

330

```
<210> 180
   <211> 41
   <212> PRT
   <213> Homo sapiens
   <220>
   <221> SITE
   <222> (41)
   <223> Xaa equals stop translation
  <400> 180
  Met Ile Ala Cys Gln Tyr Ile Ser Leu Ala Ile Met Leu Ala Phe Val
                     5
  Arg Trp Ala Ala Phe Leu Leu Phe Pro Phe Leu Cys Gly Asp Asn Gly
                20
                                    25
  Gly Asn Ile Gln Gln Lys Tyr Val Xaa
          35
                                40
  <210> 181
  <211> 52
<212> PRT
  <213> Homo sapiens
 <220>
 <221> SITE
  <222> (52)
 <223> Xaa equals stop translation
 <400> 181
 Met Ala Asn Ala Met Ala Tyr Leu Ser Ile Phe Leu Cys Gly Ala Ser
                                      10
 Ser Ser Pro Cys Asp Cys Ala Leu Leu Val Pro Val Ser Leu Phe Arg
 Gly Arg Lys Val Ala Asn Phe Lys Asn Gln Asn Ser Asp Val Thr Ser
 Gly Asn Ala Xaa
     50
<210> 182
<211> 55
<212> PRT
<213> Homo sapiens
·<220>
```

Met Gln Gln Ile Cys Ser Cys Leu Gly Ala Phe Ala Leu Leu Phe Phe Trp Pro Gly His Phe Thr Ser Thr Phe Ser Ile Phe Tyr Asp Phe Leu Pro Ile Phe Gly Ser Leu Phe Lys Cys His Pro Ser Lys Arg Pro Ser Lys Leu Pro Tyr Leu Lys Xaa 50 <210> 183 <211> 62 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (62) <223> Xaa equals stop translation <400> 183 Met Arg Leu Leu Glu Trp Arg Val Tyr Leu Arg Leu Thr Cys Ala 10 Thr Lys Asp Gly Met Ala Arg Glu Cys Pro Thr Thr Trp Leu Ser Pro 25 Pro Ala Lys Pro Asp Phe Ala Gln Arg His Ser Val Lys Pro Thr Ala 40 Leu Gln Gly Gly Arg Trp Ser Arg Leu Gly Ala Ser Pro Xaa <210> 184 <211> 148 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (148) <223> Xaa equals stop translation <400> 184 Met Leu Gly Leu Pro Trp Lys Gly Gly Leu Ser Trp Ala Leu Leu Leu Leu Leu Gly Ser Gln Ile Leu Leu Ile Tyr Ala Trp His Phe His

Glu Gln Arg Asp Cys Asp Glu His Asn Val Met Ala Arg Tyr Leu Pro
35 40 45

Ala Thr Val Glu Phe Ala Val His Thr Phe Asn Gln Gln Ser Lys Asp

```
Tyr Tyr Ala Tyr Arg Leu Gly His Ile Leu Asn Ser Trp Lys Glu Gln 65 70 75 80
```

Val Glu Ser Lys Thr Val Phe Ser Met Glu Leu Leu Gly Arg Thr 85 90 95

Arg Cys Gly Lys Phe Glu Asp Asp Ile Asp Asn Cys His Phe Gln Glu 100 105 110

Ser Thr Glu Leu Asn Asn Thr Phe Thr Cys Phe Phe Thr Ile Ser Thr 115 120 125

Arg Pro Trp Met Thr Gln Phe Ser Leu Leu Asn Lys Thr Cys Leu Glu 130 135 140

Gly Phe His Xaa 145

<210> 185

<211> 161

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (146)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (151)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (161)

<223> Xaa equals stop translation

<400> 185

Met Arg Leu Leu Cys Gly Leu Trp Leu Trp Leu Ser Leu Leu Lys Val 1 5 10 15

Leu Gln Ala Gln Thr Pro Thr Pro Leu Pro Leu Pro Pro Pro Met Gln 20 25 30

Ser Phe Gln Gly Asn Gln Phe Gln Gly Glu Trp Phe Val Leu Gly Leu 35 40 45

Ala Gly Asn Ser Phe Arg Pro Glu His Arg Ala Leu Leu Asn Ala Phe 50 55 60

Thr Ala Thr Phe Glu Leu Ser Asp Asp Gly Arg Phe Glu Val Trp Asn 65 70 75 80

Ala Met Thr Arg Gly Gln His Cys Asp Thr Trp Ser Tyr Val Leu Ile 85 90 95

Pro Ala Ala Gln Pro Gly Gln Phe Thr Val Asp His Gly Val Gly Arg 100 $$105\$

Ser Trp Leu Leu Pro Pro Gly Thr Leu Asp Gln Phe Ile Cys Leu Gly 115 120 125

Arg Ala Gln Gly Leu Ser Asp Asp Asn Ile Val Phe Pro Asp Val Thr 130 135 140

Xaa

<210> 186

<211> 122

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (122)

<223> Xaa equals stop translation

<400> 186

Met Met Leu Pro Gln Trp Leu Leu Leu Leu Phe Leu Leu Phe Phe Phe 1 5 10 15

Leu Phe Leu Leu Thr Arg Gly Ser Leu Ser Pro Thr Lys Tyr Asn Leu 20 25 30

Leu Glu Leu Lys Glu Ser Cys Ile Arg Asn Gln Asp Cys Glu Thr Gly
35 40

Cys Cys Gln Arg Ala Pro Asp Asn Cys Glu Ser His Cys Ala Glu Lys 50 55 60

Gly Ser Glu Gly Ser Leu Cys Gln Thr Gln Val Phe Phe Gly Gln Tyr
65 70 75 80

Arg Ala Cys Pro Cys Leu Arg Asn Leu Thr Cys Ile Tyr Ser Lys Asn 85 90 95

Glu Lys Trp Leu Ser Ile Ala Tyr Gly Arg Cys Gln Lys Ile Gly Arg 100 105 110

Gln Lys Leu Ala Lys Lys Met Phe Phe Xaa 115 120

<210> 187

<211> 163

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (163)

<223> Xaa equals stop translation

<400> 187

Met Thr Ser Asn Phe Pro Phe Cys Thr Leu Ile Leu Gly Ile Ala Gln
1 5 10 15

Ala Gln Ala Cys Pro Gly Cys Pro Gly Asp Trp Pro Gly Leu Gly Ser 20 25 30

Gly Val Gly Glu Gly Leu His His Ile Arg Thr Cys Arg Thr Pro Ile $35 \hspace{1cm} 40 \hspace{1cm} 45$

Pro Cys Ser Pro Pro Ala Pro Ala Ala Ala Cys Leu Gly Ser Gly His 50 55 60

Ala Arg Leu Pro Cys Val Leu Arg Leu Trp Pro Val Pro Ala Asn Leu 65 70 75 80

Ser Ser Pro Phe Arg Leu Glu Ala Leu His Cys Ser Phe Trp Ser Ser 85 90 95

Pro Leu Leu Pro Ala Pro His Leu Ala Phe Phe Gly Phe Arg Asp Leu 100 105 - 110

Leu Thr Asp Phe Leu Leu Ala Ala Cys Leu Leu Thr Phe Gln Lys Thr 115

Pro Leu Glu Leu Pro Met Ala Val Val His Leu Leu Val Ala Thr Pro 130 135 140

Cys Tyr Gln Met Leu Asp Asn Leu Pro Leu Pro Ser Ala Ala Ala Asn 150 155 150

Trp Cys Xaa

<210> 188

<211> 51

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (51)

<223> Xaa equals stop translation

<400> 188

Met Pro Gly Ile Leu Ala Gly Ile Pro Val Lys Asp Leu Cys Leu Ser 1 5 10 15

Leu Leu Gln Gly Phe Arg Leu Leu Leu Leu Cys Val Cys Pro Gly Trp

Leu Ser Gly Trp Met Gly Gly Gln Lys Gly Ser Pro Arg Ile Val Asp
35 40 45

Ile Gly Xaa

50

<222> (144)

```
<210> 189
 <211> 65
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (65)
 <223> Xaa equals stop translation
 <400> 189
 Met Tyr Leu Tyr Leu Gly Val Phe Phe His Leu Ile Tyr Pro Gly Ala
Leu Ser Ile Thr Thr Leu Gly Lys His Ser His Pro Phe Phe Thr Ala
Glu Gln Asn Ser Thr Val Trp Met Glu His Thr Leu Phe His Gln Ser
Pro Val Ala Ser His Leu Val Cys Phe Gln Ser Phe Ala Phe Ser Glu
Xaa
 65
<210> 190
<211> 47
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (47)
<223> Xaa equals stop translation
<400> 190
Met Thr Leu Ser Leu Gln Leu Ala Glu Leu Val His Phe Val Cys Ala
                                                         15
Phe Gln Ser Gln Trp Thr Gly Val Tyr Pro Met Met Pro Pro Leu Lys
Pro Thr Glu Pro Leu Cys Phe Ala Cys Val Pro Cys Arg Val Xaa
                             40
<210> 191
<211> 144
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
```

<223> Xaa equals stop translation

<400> 191

Met Ser Pro Phe His Leu Leu Gly Leu Lys Val Phe Leu Thr Trp Ala 1 5 10 15

Leu Thr Leu Ala Gln Ile Cys Leu Tyr Phe Phe Glu Val Gln Pro Leu 20 25 30

Gly Leu Leu Ala Leu Asn Phe Phe Cys Thr Ala Thr Ala Gly Leu Lys
35 . 40 . 45

Glu Leu Cys Met His Pro Pro Ser Leu Ala Phe Thr Pro Glu Phe His 50 55 60

Thr Ser Leu Ser Pro Leu Ala Ile Pro Ser Phe Cys Gly Thr Ser Val 65 70 75 80

Ser Leu Ser Asn Ser His Thr Ile Pro Leu Ser Leu Tyr Leu Pro Phe 85 90 95

Leu Pro Leu Val His Ser Gln Val Leu Pro Val Lys Asp Val Thr Ile

Glu Trp Pro Leu Cys Gln Arg Cys Leu Gly Ser Thr Cys His Gln Xaa 130 135 140

<210> 192

<211> 81

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (76)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (81)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 192

Met Phe Cys Phe Ser Ser Ile Phe Cys Ser His Glu His Thr His Leu

1 10 15

Pro Gly Thr Phe Trp Leu Phe Leu Phe Leu Phe Leu Ile Leu Pro Pro 20 30

Ser Cys Pro Cys Phe Leu Pro Phe Ser Leu Ala Ile Glu Thr Val Arg

```
Trp Pro Cys Trp His His Pro Thr Ser Phe Glu Leu Cys Tyr Pro Gly 50 60
```

Thr Ser Ile Tyr Tyr Ala Ser Arg Gly Gly Pro Xaa Pro Asn Ser Glu 65 70 75 80

Xaa

<210> 193 <211> 45

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (45)

<223> Xaa equals stop translation

<400> 193

Met Thr Tyr Leu Phe Cys Ser Ser Ile Ser Leu Leu Leu Leu Lys Val 1 5 10 15

His Ser Ser Gly His Gln Asp Ile Arg Lys Ala Lys Ser Lys Val Pro $20 \hspace{1cm} 25 \hspace{1cm} 30$

Arg Leu Leu Ile Ile Gln Cys Pro Gln Gln Arg Glu Xaa 35 40 45

<210> 194

<211> 42

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (42)

<223> Xaa equals stop translation

<400> 194

Met Pro Thr Ile Trp Val Lys Leu Cys Leu Leu Gln Val Cys His Gly $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Leu Phe Pro Leu Leu Lys His Trp Ser Gln Pro Met Pro Leu Cys Val

Thr Leu Ala Pro Val Ser Tyr Trp Leu Xaa

<210> 195

<211> 260

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (260)

<223> Xaa equals stop translation

<400> 195

Met Gly Thr Ala Ala Leu Gly Pro Val Trp Ala Ala Leu Leu Leu Phe 1 5 10 15

Leu Leu Met Cys Glu Ile Pro Met Val Glu Leu Thr Phe Asp Arg Ala 20 25 30

Val Ala Ser Asp Cys Gln Arg Cys Cys Asp Ser Glu Asp Pro Leu Asp 35 40 45

Pro Ala His Val Ser Ser Ala Ser Ser Ser Gly Arg Pro His Ala Leu
50 55 60

Pro Glu Ile Arg Pro Tyr Ile Asn Ile Thr Ile Leu Lys Gly Asp Lys 65 70 75 80

Gly Asp Pro Gly Pro Met Gly Leu Pro Gly Tyr Met Gly Arg Glu Gly 85 90 95

Pro Gln Gly Glu Pro Gly Pro Gln Gly Ser Lys Gly Asp Lys Gly Glu 100 105 110

Met Gly Ser Pro Gly Ala Pro Cys Gln Lys Arg Phe Phe Ala Phe Ser 115 120 125

Val Gly Arg Lys Thr Ala Leu His Ser Gly Glu Asp Phe Gln Thr Leu 130 135 140

Leu Phe Glu Arg Val Phe Val Asn Leu Asp Gly Cys Phe Asp Met Ala 145 150 155 160

Thr Gly Gln Phe Ala Ala Pro Leu Arg Gly Ile Tyr Phe Phe Ser Leu 175 to 175

Asn Val His Ser Trp Asn Tyr Lys Glu Thr Tyr Val His Ile Met His 180 185 190

Asn Gln Lys Glu Ala Val Ile Leu Tyr Ala Gln Pro Ser Glu Arg Ser 195 200 205

Ile Met Gln Ser Gln Ser Val Met Leu Asp Leu Ala Tyr Gly Asp Arg 210 215 220

Val Trp Val Arg Leu Phe Lys Arg Gln Arg Glu Asn Ala Ile Tyr Ser 225 230 235 240

Asn Asp Phe Asp Thr Tyr Ile Thr Phe Ser Gly His Leu Ile Lys Ala 245 250 255

Glu Asp Asp Xaa 260

<210> 196

<211> 117

<212> PRT

<213> Homo sapiens

<400> 196

Met Leu Gly His Cys Cys Tyr Phe Trp Gln Val Trp Pro Ala Ser Glu
1 5 10 15

Ala Leu Ala Ala Gly Pro Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser 20 25 30

Trp Lys Gln His Ile Gly Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu $35 \hspace{1cm} 40 \hspace{1cm} 45$

Pro Thr Thr Leu Thr Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly 50 60

Lys Asn Pro Ala Ala Gly Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly 65 70 75 80

Val Trp Pro Cys Phe Ala Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr 85 90 95

Pro Ser Ser Thr Gly Leu Arg Ser Cys Leu Val Arg Ser Pro Ala Thr 100 105 110

Trp Trp Arg Thr Pro 115

<210> 197

<211> 698

<212> PRT

<213> Homo sapiens

<400> 197

Met Leu Pro Ala Arg Leu Pro Phe Arg Leu Leu Ser Leu Phe Leu Arg

1 10 15

Gly Ser Ala Pro Thr Ala Ala Arg His Gly Leu Arg Glu Pro Leu Leu 20 25 30

Glu Arg Arg Cys Ala Ala Ala Ser Ser Phe Gln His Ser Ser Leu 35 40 45

Gly Arg Glu Leu Pro Tyr Asp Pro Val Asp Thr Glu Gly Phe Gly Glu 50 55 60

Gly Gly Asp Met Gln Glu Arg Phe Leu Phe Pro Glu Tyr Ile Leu Asp 65 70 75 80

Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Gln 85 90 95

Gln Gln Glu Glu Glu Arg Gln Arg Gln Gln Arg Glu Glu Arg 100 105 110

Arg Gln Gln Asn Leu Arg Ala Arg Ser Arg Glu His Pro Val Val Gly
115 120 125

His Pro Asp Pro Ala Leu Pro Pro Ser Gly Val Asn Cys Ser Gly Cys

	130					135					1	L40				
Gly . 145	Ala (Glu I	Leu I	dis (Cys (Gln	Asp	Al.	a G	ly V 1	al E 55	Pro	Gly	Ту	Le	u Pr 16
Arg (Glu I	ys F	he I	eu A .65	rg T	Chr .	Ala	Glı	ı A.	la A 70	sp G	ly (Gly	Let	1 Al.	
Thr V	/al C	ys G 1	ln A 80	rg C	ys I	rp :	Leu	Let 185	ı Se	er H	is H	is A	Arg	Arg 190		a Le
Arg I	eu G 1	ln V 95	al S	er A	rg G	lu (Gln 200	Туг	· Le	u G	lu L	eu (2	7a1 205	Ser	Ala	a Ala
Leu A 2	rg A	rg P	ro G	ly P	ro S 2	er I 15	Leu	Val	Le	u Ty	r Me 22	et V 20	al	Asp	Leu	Let
Asp L 225				۷.						23	5					240
Lys G									250	J					255	
Ala P							•	265					2	270		
Ala Aı						4	60					28	35			
Pro Va 29	_				49	5					30	D				
Trp Se				31	,					315						320
Gly Ty			J2.	,					330					3	335	
Tyr Ar							٠	45					3	50		
Thr Le						20	U					369	5			
Ser Ası 370	Ala	Ile	Asp	Arg	Ala 375	Th:	r II	le S	Ser	Pro	Trp 380	Pro	G]	Ly T	hr (Thr
Leu Asr 385	l Leu	Leu	Lys	Phe 390	Pro	Ile	e C ⁷	/S A	sn	Pro 395	Thr	Pro	Ту	r A		iet 100
Phe Lys	Arg	His	Gln 405	Arg	Leu	Lys	E Ly	rs A 4	sp 10	Ser	Thr	Gln	Al	a G 4	lu G 15	lu
Asp Leu	Ser	Glu 420	Gln	Glu	Gln	Asn	Gl 42	n L 5	eu .	Asn	Val	Leu	Ly.		/s H	is
Gly Tyr	Val 435	Val	Gly	Arg	Val	Gly 440	Ar	g T	hr 1	Phe	Leu	Tyr 445	Se	r Gl	lu G	lu

- Gln Lys Asp Asn Ile Pro Phe Glu Phe Asp Ala Asp Ser Leu Ala Phe 450 455 460
- Asp Met Glu Asn Asp Pro Val Met Gly Thr His Lys Ser Thr Lys Gln 465 470 475 480
- Val Glu Leu Thr Ala Gln Asp Val Lys Asp Ala His Trp Phe Tyr Asp
 485 490 495
- Thr Pro Gly Ile Thr Lys Glu Asn Cys Ile Leu Asn Leu Leu Thr Glu 500 505 510
- Lys Glu Val Asn Ile Val Leu Pro Thr Gln Ser Ile Val Pro Arg Thr 515 520 525
- Phe Val Leu Lys Pro Gly Met Val Leu Phe Leu Gly Ala Ile Gly Arg 530 540
- Ile Asp Phe Leu Gln Gly Asn Gln Ser Ala Trp Phe Thr Val Val Ala 545 550 555 560
- Ser Asn Ile Leu Pro Val His Ile Thr Ser Leu Asp Arg Ala Asp Ala 575
- Leu Tyr Gln Lys His Ala Gly His Thr Leu Leu Gln Ile Pro Met Gly 580 585 590
- Gly Lys Glu Arg Met Ala Gly Phe Pro Pro Leu Val Ala Glu Asp Ile 595 600 605
- Met Leu Lys Glu Gly Leu Gly Ala Ser Glu Ala Val Ala Asp Ile Lys 610 615 620
- Phe Ser Ser Ala Gly Trp Val Ser Val Thr Pro Asn Phe Lys Asp Arg 625 630 635 640
- Leu His Leu Arg Gly Tyr Thr Pro Glu Gly Thr Val Leu Thr Val Arg
 645 650 655
- Pro Pro Leu Leu Pro Tyr Ile Val Asn Ile Lys Gly Gln Arg Ile Lys 660 665 670
- Lys Ser Val Ala Tyr Lys Thr Lys Lys Pro Pro Ser Leu Met Tyr Asn 675 680 685
- Val Arg Lys Lys Gly Lys Ile Asn Val 695

<210> 198

<211> 348

<212> PRT

<213> Homo sapiens

<400> 198

Met Asn Met Thr Gln Ala Arg Val Leu Val Ala Ala Val Val Gly Leu

1 5 10 15

Val Ala Val Leu Leu Tyr Ala Ser Ile His Lys Ile Glu Glu Gly His

20		25	30
Leu Ala Val Thre	There's are all		
		40	Thr Ser Pro Ser Gly
Pro Gly Tyr His 50	Ile Met Leu 55	Pro Phe Ile Thr	Thr Phe Arg Ser Val
Gln Thr Thr Leu 65	Gln Thr Asp	Glu Val Lys Asn 75	Val Pro Cys Gly Thr
Ser Gly Gly Val	Met Ile Tyr :	Ile Asp Arg Ile	Glu Val Val Asn Met
	85	90	95
Leu Ala Pro Tyr 100	Ala Val Phe A	Asp Ile Val Arg	Asn Tyr Thr Ala Asp
Tyr Asp Lys Thr 1	Leu Ile Phe <i>I</i>	asn Lys Ile His	His Glu Leu Asn Gln
Phe Cys Ser Ala F 130	His Thr Leu G 135	ln Glu Val Tyr	Ile Glu Leu Phe Asp
Gln Ile Asp Glu A	asn Leu Lys G	ln Ala Leu Gln 1	Lys Asp Leu Asn Leu
145	150	155	160
Met Ala Pro Gly L	eu Thr Ile G	ln Ala Val Arg v	Val Thr Lys Pro Lys
1	65	170	
Ile Pro Glu Ala I	le Arg Arg A	sn Phe Glu Leu M	Met Glu Ala Glu Lys
180		185	190
Thr Lys Leu Leu I	le Ala Ala Gl	n Lys Gln Lys V	al Val Glu Lys Glu
195	20		205
Ala Glu Thr Glu An	rg Lys Lys Al	a Val Ile Glu A	la Glu Lys Ile Ala
210	215		20
Gln Val Ala Lys II	le Arg Phe Gl	n Gln Lys Val M	et Glu Lys Glu Thr
225	230	235	240
Glu Lys Arg Ile Se	er Glu Ile Gl	u Asp Ala Ala Pl	·
24	5	250	
Lys Ala Lys Ala As	p Ala Glu Ty:	r Tyr Ala Ala Hi	s Lys Tyr Ala Thr
260		265	270
Ser Asn Lys His Ly	s Leu Thr Pro	o Glu Tyr Leu Gl	u Leu Lys Lys Tyr
275)	285
Gln Ala Ile Ala Se	r Asn Ser Lys	: Ile Tyr Phe Gl	y Ser Asn Ile Pro
290	295	30	0
Asn Met Phe Val Asp	Ser Ser Cys	Ala Leu Lys Ty	r Ser Asp Ile Arg
305	310	315	320
Thr Gly Arg Glu Ser	Ser Leu Pro	Ser Lys Glu Ala 330	a Leu Glu Pro Ser 335

Gly Glu Asn Val Ile Gln Asn Lys Glu Ser Thr Gly 340 345

<210> 199

<211> 401 <212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (307)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 199

Met Met Gly Leu Gly Asn Gly Arg Arg Ser Met Lys Ser Pro Pro Leu

1 10 15

Val Leu Ala Ala Leu Val Ala Cys Ile Ile Val Leu Gly Phe Asn Tyr
20 25 30

Trp Ile Ala Ser Ser Arg Ser Val Asp Leu Gln Thr Arg Ile Met Glu 35 40 45

Leu Glu Gly Arg Val Arg Arg Arg Ala Ala Glu Arg Gly Ala Val Glu 50 55 60

Leu Lys Lys Asn Glu Phe Gln Gly Glu Leu Glu Lys Gln Arg Glu Gln 65 70 75 80

Leu Asp Lys Ile Gln Ser Ser His Asn Phe Gln Leu Glu Ser Val Asn 85 90 95

Lys Leu Tyr Gln Asp Glu Lys Ala Val Leu Val Asn Asn Ile Thr Thr 100 105 110

Gly Glu Arg Leu Ile Arg Val Leu Gln Asp Gln Leu Lys Thr Leu Gln 115 120 125

Arg Asn Tyr Gly Arg Leu Gln Gln Asp Val Leu Gln Phe Gln Lys Asn 130 135 140

Gln Thr Asn Leu Glu Arg Lys Phe Ser Tyr Asp Leu Ser Gln Cys Ile 145 150 155 160

Asn Gln Met Lys Glu Val Lys Glu Gln Cys Glu Glu Arg Ile Glu Glu 165 170 175

Val Thr Lys Lys Gly Asn Glu Ala Val Ala Ser Arg Asp Leu Ser Glu 180 185 190

Asn Asn Asp Gln Arg Gln Gln Leu Gln Ala Leu Ser Glu Pro Gln Pro 195 200 205

Arg Leu Gln Ala Ala Gly Leu Pro His Thr Glu Val Pro Gln Gly Lys 210 225

Gly Asn Val Leu Gly Asn Ser Lys Ser Gln Thr Pro Ala Pro Ser Ser 225 230 235 240

Glu Val Val Leu Asp Ser Lys Arg Gln Val Glu Lys Glu Glu Thr Asn 245 250 255

Glu Ile Gln Val Val Asn Glu Glu Pro Gln Arg Asp Arg Leu Pro Gln 260 265 270

Glu Pro Gly Arg Glu Gln Val Val Glu Asp Arg Pro Val Gly Gly Arg 275 280 285

Gly Phe Gly Gly Ala Gly Glu Leu Gly Gln Thr Pro Gln Val Gln Ala 290 295 300

Ala Leu Xaa Val Ser Gln Glu Asn Pro Glu Met Glu Gly Pro Glu Arg

Asp Gln Leu Val Ile Pro Asp Gly Gln Glu Glu Glu Gln Glu Ala Ala 325 330 335

Gly Glu Gly Arg Asn Gln Gln Lys Leu Arg Gly Glu Asp Asp Tyr Asn 340 345 350

Met Asp Glu Asn Glu Ala Glu Ser Glu Thr Asp Lys Gln Ala Ala Leu 355 360 365

Ala Gly Asn Asp Arg Asn Ile Asp Val Phe Asn Val Glu Asp Gln Lys 370 380

Arg Asp Thr Ile Asn Leu Leu Asp Gln Arg Glu Lys Arg Asn His Thr 385 395 400

Leu

<210> 200

<211> 324

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (3)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 200

Met Glu Xaa Ala Lys Val Tyr Val Ala Lys Val Asp Cys Thr Ala His

1 10 15

Ser Asp Val Cys Ser Ala Gln Gly Val Arg Gly Tyr Pro Thr Leu Lys

Leu Phe Lys Pro Gly Gln Glu Ala Val Lys Tyr Gln Gly Pro Arg Asp 35 40 45

Phe Gln Thr Leu Glu Asn Trp Met Leu Gln Thr Leu Asn Glu Glu Pro
50 55 60

Val Thr Pro Glu Pro Glu Val Glu Pro Pro Ser Ala Pro Glu Leu Lys

6.	5				7	0				7	75				80
Gli	n Gl	y Le	u Ty	r Gl 8	u Lei 5	u Se	r Al	a Se	r As	n Ph	e Gl	u Le	u His	Va.	
Glr	ı Gl	y As	р Ні 10	s Ph 0	e Ile	e Ly	s Ph	e Ph 10	e Al 5	a Pr	o Tr	Cy:	s Gly 110		Cys
Lys	Ala	a Le 11	u Al 5	a Pr	o Thr	Tr	p Gl	u Gl 0 ·	n Le	u Al	a Lei	Gly 125	/ Leu 5	Glu	His
Ser	Gli 130	ıTh:	r Va	l Lys	s Ile	Gl ₃	y Ly: 5	s Va	l As	p Cy:	s Thr 140		His	Tyr	Glu
Leu 145	Cys	S Se:	r Gl	y Asr	1 Gln 150	Val	l Arg	g Gl	ү Ту:	r Pro	o Thr	Leu	Leu	Trp	Phe 160
Arg	Asp	Gly	/ Lys	165	Val	Asp	Glr	ту:	170	s Gly	/ Lys	Arg	Asp	Leu 175	Glu
Ser	Leu	Arc	180	ı Tyr	Val	Glu	Ser	Glr 185	Let	ı Glr	a Arg	Thr	Glu 190	Thr	Gly
Ala	Thr	Glu 195	Thr	· Val	Thr	Pro	Ser 200	Glu	Ala	Pro	Val	Leu 205	Ala	Ala	Glu
Pro	Glu 210	Ala	Asp	Lys	Gly	Thr 215	Val	Leu	Ala	Leu	Thr 220	Glu	Asn	Asn	Phe
Asp 225	Asp	Thr	Ile	Ala	Glu 230	Gly	Ile	Thr	Phe	Ile 235	Lys	Phe	Tyr	Ala	Pro 240
Trp	Cys	Gly	His	Cys 245	Lys	Thr	Leu	Ala	Pro 250	Thr	Trp	Glu	Glu	Leu 255	Ser
Lys	Lys	Glu	Phe 260	Pro	Gly	Leu	Ala	Gly 265	Val	Lys	Ile	Ala	Glu 270	Val	Asp
Cys	Thr	Ala 275	Glu	Arg	Asn	Ile	Cys 280	Ser	Lys	Tyr		Val 285	Arg	Gly '	Tyr
Pro	Thr 290	Leu	Leu	Leu	Phe .	Arg 2 9 5	Gly	Gly	Lys	Lys	Val 300	Ser	Glu I	lis :	Ser
Gly 6 305	Gly	Arg	Asp	Leu	Asp :	Ser	Leu	His	Arg	Phe 315	Val :	Leu	Ser (Ala 320
Lys 2	Asp	Glu	Leu												
<211><212>	<210> 201 <211> 90 <212> PRT <213> Homo sapiens														
<400>	20:	1													

Met Ala Leu Phe Ser Cys Leu Leu Leu Leu Lys Gln Ser Asp Gly Ala

1 5 10 15

- Ser Pro Val Leu Arg Ala Leu Ala Ala Ser Cys Leu Ala Ser Pro Ala 20 25 30
- Gly Cys Cys Gly Thr Arg Lys Ala Leu Asn Gly Asn Val Gly Glu Lys 35 40 45
- Val Gly Phe Thr Phe Met Ser Phe Gln Gly Cys Asp Pro Ser Ser Pro 50 55 60
- Gly Cys Leu Cys Cys Ser Leu Leu Pro Ser Asn Ser Gln Leu Val Phe
 65 70 75 80
- Ile Ser Phe Leu Val Leu Ser Gly Leu Ala 85 90
- <210> 202
- <211> 243
- <212> PRT
- <213> Homo sapiens
- <400> 202
- Met Arg Pro Gln Gly Pro Ala Ala Ser Pro Gln Arg Leu Arg Gly Leu

 1 10 15
- Leu Leu Leu Leu Leu Gln Leu Pro Ala Pro Ser Ser Ala Ser Glu
 20 25 30
- Ile Pro Lys Gly Lys Gln Lys Ala Gln Leu Arg Gln Arg Glu Val Val 35 40 45
- Asp Leu Tyr Asn Gly Met Cys Leu Gln Gly Pro Ala Gly Val Pro Gly 50 55 60
- Arg Asp Gly Ser Pro Gly Ala Asn Gly Ile Pro Gly Thr Pro Gly Ile
 65 70 75 80
- Pro Gly Arg Asp Gly Phe Lys Gly Glu Lys Gly Glu Cys Leu Arg Glu 85 90 95
- Ser Phe Glu Glu Ser Trp Thr Pro Asn Tyr Lys Gln Cys Ser Trp Ser 100 105 110
- Ser Leu Asn Tyr Gly Ile Asp Leu Gly Lys Ile Ala Glu Cys Thr Phe 115 120 125
- Thr Lys Met Arg Ser Asn Ser Ala Leu Arg Val Leu Phe Ser Gly Ser 130 140
- Leu Arg Leu Lys Cys Arg Asn Ala Cys Cys Glu Arg Trp Tyr Phe Thr 155 150 160
- Phe Asn Gly Ala Glu Cys Ser Gly Pro Leu Pro Ile Glu Ala Ile Ile 165 - 170 - 175
- Tyr Leu Asp Gln Gly Ser Pro Glu Met Asn Ser Thr Ile Asn Ile His 180 185 190

WO 00/06698 PCT/US99/17130

145

Arg Thr Ser Ser Val Glu Gly Leu Cys Glu Gly Ile Gly Ala Gly Leu 195 200 205

Val Asp Val Ala Ile Trp Val Gly Thr Cys Ser Asp Tyr Pro Lys Gly 210 215 220

Asp Ala Ser Thr Gly Trp Asn Ser Val Ser Arg Ile Ile Glu Glu 225 230 235 240

Leu Pro Lys

<210> 203

<211> 75

<212> PRT

<213> Homo sapiens

<400> 203

Met Ala Gly Gln Glu Asp Pro Val Gln Arg Glu Ile His Gln Asp Trp

1 5 10 15

Ala Asn Arg Glu Tyr Ile Glu Ile Ile Thr Ser Ser Ile Lys Lys Ile 20 25 30

Ala Asp Phe Leu Asn Ser Phe Asp Met Ser Cys Arg Ser Arg Leu Ala 35 40 45

Thr Leu Asn Glu Lys Leu Thr Ala Leu Glu Arg Arg Ile Glu Tyr Ile
50 60

Glu Ala Arg Val Thr Lys Gly Glu Thr Leu Thr 65 70 75

<210> 204

<211> 248

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (185)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 204

Met Thr Ser Gln Pro Val Pro Asn Glu Thr Ile Ile Val Leu Pro Ser

Asn Val Ile Asn Phe Ser Gln Ala Glu Lys Pro Glu Pro Thr Asn Gln
20 25 30

Gly Gln Asp Ser Leu Lys Lys His Leu His Ala Glu Ile Lys Val Ile 35 40 45

Gly Thr Ile Gln Ile Leu Cys Gly Met Met Val Leu Ser Leu Gly Ile
50 60

Ile Leu Ala Ser Ala Ser Phe Ser Pro Asn Phe Thr Gln Val Thr Ser

80

6	5				70)				75	i				80
Thi	r Lei	u Le	u Ası	n Ser 85	Ala	Туг	Pro	Phe	: Ile 90	Gly	Pro	Phe	. Phe	Phe 95	
Ile	e Sei	r Gly	y Ser 100	r Leu	Ser	lle	Ala	Thr	Glu	Lys	Arg	Leu	Thr		Leu
Leu	ı Val	His 115	Ser	Ser	Leu	Val	Gly 120	Ser	Ile	Leu	Ser	Ala 125		Ser	Ala
Leu	Val 130	Gly	' Phe	: Ile	Ile	Leu 135	Ser	Val	Lys	Gln	Ala 140	Thr	Leu	Asn	Pro
Ala 145	Ser	Leu	Gln	Cys	Glu 150	Leu	Asp	Lys	Asn	Asn 155	Ile	Pro	Thr	Arg	Ser 160
Tyr	Val	Ser	Tyr	Phe 165	Tyr	His	Asp	Ser	Leu 170	Tyr	Thr	Thr	Asp	Cys 175	Tyr
Thr	Ala	Lys	Ala 180	Ser	Leu	Ala	Gly	Xaa 185	Leu	Ser	Leu	Met	Leu 190	Ile	Cys
Thr	Leu	Leu 1'95	Glu	Phe	Cys	Leu	Ala 200	Val	Leu	Thr		Val 205	Leu	Arg	Trp
Lys	Gln 210	Ala	Tyr	Ser	Asp	Phe 215	Pro	Gly	Ser	Val	Leu 220	Phe	Leu	Pro	His
Ser 225	Tyr	Ile	Gly	Asn	Ser 230	Gly	Met	Ser	Ser	Lys 1 235	Met '	Thr	His .		Cys 240
Gly	Tyr	Glu		Leu : 245	Leu '	Thr	Ser		•		,	•			
<210 <211 <212 <213	> 16 > PR	8 T	apie:	ns											
<220 <221 <222 <223	> > SI' > (8)	TE 3)			of t	he n	ıatur	allv	occ	urri	na I.	-ami	no a	cido	
<400	> 20	5													
Met 1				Э					10					15	
Pro I	Leu F	lis T	Thr G 20	ilu A	la V	al V	al L	eu L 25	eu V	al P	ro S		sp A 30	sp G	ly
Arg A	la F	he L 35	eu L	eu A	rg S	er A	rg Lo 40	eu L	eu H	is Pı		lu A. 15	la H	is V	al
Pro P	ro A 50	la A	la A	sp A	rg G	ly A:	la Se	er Le	eu G		ys Va 50	al Le	eu H:	is G	ln

Ala Ala Pro Lys Ser Arg Pro Arg Ser Pro Ala Ala Gly Ala Ala Leu 65 70 75 80

Leu His Xaa Pro Arg Arg Thr Gly Asp Glu Pro Cys Arg Glu Phe His 85 90 95

Gly Asn Gly Phe Pro Gly Pro Thr Gln Leu Thr Pro Gly Glu Cys Gly
100 105 110

Leu Pro Ala Pro Ser Ser Leu Leu Gln His Ala Ser Ala Pro Val Arg 115 120 125

Thr Gly Ser Glu Gly Gln Val Val Gly Cys Pro Arg Ala Arg Gly Glu 130 135 140

Thr Gly Glu Gly Leu Ser Leu Ala Phe Leu Ser Ser Leu Met Phe Thr 145 150 155 160

Ser Arg Asn Gly Leu Val Gly Cys 165

<210> 206

<211> 218

<212> PRT

<213> Homo sapiens

<400> 206

Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val

Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val 20 25 30

Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Trp Lys

Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly His Met Gln Cys 50 60

Lys Val Tyr Asp Ser Val Leu Ala Leu Ser Thr Glu Val Gln Ala Ala 65 70 75 80

Arg Ala Leu Thr Val Ser Ala Val Leu Leu Ala Phe Val Ala Leu Phe
85 90 95

Val Thr Leu Ala Gly Ala Gln Cys Thr Thr Cys Val Ala Pro Gly Pro 100 105 110

Ala Lys Ala Arg Val Ala Leu Thr Gly Gly Val Leu Tyr Leu Phe Cys 115 120 125

Gly Leu Leu Ala Leu Val Pro Leu Cys Trp Phe Ala Asn Ile Val Val 130 135 140

Arg Glu Phe Tyr Asp Pro Ser Val Pro Val Ser Gln Lys Tyr Glu Leu 145 150 155 160

Gly Ala Ala Leu Tyr Ile Gly Trp Ala Ala Thr Ala Leu Leu Met Val

165 170 175

Gly Gly Cys Leu Leu Cys Cys Gly Ala Trp Val Cys Thr Gly Arg Pro 180 185 190

Asp Leu Ser Phe Pro Val Lys Tyr Ser Ala Pro Arg Arg Pro Thr Ala 195 200 205

Thr Gly Asp Tyr Asp Lys Lys Asn Tyr Val 210 215

<210> 207

<211> 73

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (73)

<223> Xaa equals stop translation

<400> 207

Met Thr Ser Tyr Ile Leu Ile Ser Phe Val Leu Leu Ile Gly Val Gly

1 5 10 15

Cys Ile Glu Lys Asp Gln Ser Cys Pro Val Phe Gly Gly Arg Lys Arg 20 25 30

Leu His Leu Leu Phe Val Gly Gly Gln Leu Arg Gln Val Arg Met Leu 35 40 45

Arg Gly Glu Leu Ser Cys Ala Cys Tyr Arg Pro His Val Gln Ala Leu 50 55 60

Gln Leu Gly Gly Cys Thr Cys Phe Xaa 65 70

<210> 208

<211> 348

<212> PRT

<213> Homo sapiens

<400> 208

Met Leu Cys Pro Trp Arg Thr Ala Asn Leu Gly Leu Leu Leu Ile Leu 1 5 10 15

Thr Ile Phe Leu Val Ala Glu Ala Glu Gly Ala Ala Gln Pro Asn Asn 20 25 30

Ser Leu Met Leu Gln Thr Ser Lys Glu Asn His Ala Leu Ala Ser Ser

Ser Leu Cys Met Asp Glu Lys Gln Ile Thr Gln Asn Tyr Ser Lys Val 50 55 60

Leu Ala Glu Val Asn Thr Ser Trp Pro Val Lys Met Ala Thr Asn Ala 65 70 75 80

Val Leu Cys Cys Pro Pro Ile Ala Leu Arg Asn Leu Ile Ile Ile Thr 85 90 95

Trp Glu Ile Ile Leu Arg Gly Gln Pro Ser Cys Thr Lys Ala Tyr Lys
100 105 110

Lys Glu Thr Asn Glu Thr Lys Glu Thr Asn Cys Thr Asp Glu Arg Ile 115 120 125

Thr Trp Val Ser Arg Pro Asp Gln Asn Ser Asp Leu Gln Ile Arg Thr 130 140

Val Ala Ile Thr His Asp Gly Tyr Tyr Arg Cys Ile Met Val Thr Pro 145 150 155 160

Asp Gly Asn Phe His Arg Gly Tyr His Leu Gln Val Leu Val Thr Pro 165 170 175

Glu Val Thr Leu Phe Gln Asn Arg Asn Arg Thr Ala Val Cys Lys Ala 180 185 190

Val Ala Gly Lys Pro Ala Ala His Ile Ser Trp Ile Pro Glu Gly Asp 195 200 205

Cys Ala Thr Lys Gln Glu Tyr Trp Ser Asn Gly Thr Val Thr Val Lys 210 215 220

Ser Thr Cys His Trp Glu Val His Asn Val Ser Thr Val Asn Cys His 225 230 235 240

Val Ser His Leu Thr Gly Asn Lys Ser Leu Tyr Ile Glu Leu Leu Pro 245 250 255

Val Pro Gly Ala Lys Lys Ser Ala Lys Leu Tyr Ile Pro Tyr Ile Ile 260 265 270

Leu Thr Ile Ile Ile Leu Thr Ile Val Gly Phe Ile Trp Leu Leu Lys 275 280 285

Val Asn Gly Cys Arg Lys Tyr Lys Leu Asn Lys Thr Glu Ser Thr Pro 290 295 300

Val Val Glu Glu Asp Glu Met Gln Pro Tyr Ala Ser Tyr Thr Glu Lys 305 310 315 320

Asn Asn Pro Leu Tyr Asp Thr Thr Asn Lys Val Lys Ala Ser Glu Ala 325 330 335

Leu Gln Ser Glu Val Asp Thr Asp Leu His Thr Leu 340 345

<210> 209

<211> 73

<212> PRT

<213> Homo sapiens

<220>

```
<221> SITE
```

<222> (73)

<223> Xaa equals stop translation

<400> 209

Met Ala Arg Gly Cys Val Cys Ser Leu Cys Ala Ser Val Cys Ile Phe 1 5 10 15

Leu Ser Ser Leu Phe Pro Leu Leu Pro Ser Val His Ser Val Asn Ile
20 25 30

Ile Ser Cys Leu Leu Leu Ser Lys Cys Phe Glu Gly Leu Glu Leu Met
35 40

Cys Glu His Leu Tyr Gln Leu Ser Gln Leu His Val Leu His His Ile 50 55 60

Phe Ser Tyr Leu Leu Cys Thr Pro Xaa 65 70

<210> 210

<211> 608

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (265)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (597)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 210

Met Val Gly Thr Lys Leu Arg Gln Thr Lys Asp Ala Leu Phe Thr Ile
1 5 10 15

Leu His Asp Leu Arg Pro Gln Asp Arg Phe Ser Ile Ile Gly Phe Ser 20 25 30

Asn Arg Ile Lys Val Trp Lys Asp His Leu Ile Ser Val Thr Pro Asp 35 40 45

Ser Ile Arg Asp Gly Lys Val Tyr Ile His His Met Ser Pro Thr Gly
50 55 60

Gly Thr Asp Ile Asn Gly Val Leu Gln Arg Ala Ile Arg Leu Leu Asn 65 70 75 80

Lys Tyr Val Ala His Ser Gly Ile Gly Asp Arg Ser Val Ser Leu Ile 85 90 95

Val Phe Leu Thr Asp Gly Lys Pro Thr Val Gly Glu Thr His Thr Leu
100 105 110

Lys Ile Leu Asn Asn Thr Arg Glu Ala Ala Arg Gly Gln Val Cys Ile

		115	5				120	כ				125	5		
Ph€	Thr 130		≘ Gly	/ Ile	e Gly	/ Asr 135		o Val	l Asp	Phe	Arg 140		ı Lev	ı Glı	Lys
Leu 145		: Le	ı Glu	ı Asr	150		/ Let	ı Thr	: Arg	155		. His	s Glu	Glu	Glu 160
Asp	Ala	Gly	/ Ser	Glr 165		Ile	e Gly	/ Phe	туr 170		Glu	ı Ile	e Arg	Thr 175	
Leu	Leu	Ser	180		Arg	Ile	Asp	185		Pro	Ser	Ser	Val 190		Gln
Ala	Thr	Lys 195		Leu	Phe	Pro	Asn 200		Phe	Asn	Gly	Ser 205		Ile	Ile
Ile	Ala 210		Lys	Leu	Val	Asp 215		Lys	Leu	Asp	His 220		His	Val	Glu
Val 225		Ala	Ser	Asn	Ser 230	Lys	Lys	Phe	Ile	Ile 235	Leu	Lys	Thr	Asp	Val 240
Pro	Val	Arg	Pro	Gln 245	Lys	Ala	Gly	Lys	Asp 250	Val	Thr	Gly	Ser	Pro 255	Arg
Pro	Gly	Gly	Asp 260	Gly	Glu	Gly	Asp	Xaa 265	Asn	His	Ile	Glu	Arg 270	Leu	Trp
Ser	Tyr	Leu 275		Thr	Lys	Glu	Leu 280	Leu	Ser	Ser	Trp	Leu 285	Gln	Ser	Asp
Asp	Glu 290	Pro	Glu	Lys	Glu	Arg 295	Leu	Arg	Gln	Arg	Ala 300	Gln	Ala	Leu	Ala
Val 305	Ser	Tyr	Arg	Phe	Leu 310	Thr	Pro	Phe	Thr	Ser 315	Met	Lys	Leu	Arg	Gly 320
Pro	Val	Pro	Arg	Met 325	Asp	Gly	Leu	Glu	Glu 330	Ala	His	Gly	Met	Ser 335	Ala
Ala	Met	Gly	Pro 340	Glu	Pro	Val	Val	Gln 345	Ser	Val	Arg	Gly	Ala 350	Gly	Thr
Gln	Pro	Gly 355	Pro	Leu	Leu	Lys	Lys 360	Pro	Tyr	Gln	Pro	Arg 365	Ile	Lys	Ile
Ser	Lys 370	Thr	Ser	Val	Asp	Gly 375	Asp	Pro	His	Phe	Val 380	Val	Asp	Phe	Pro
Leu 385	Ser	Arg	Leu	Thr	Val 390	Cys	Phe	Asn	Ile	Asp 395	Gly	Gln	Pro		Asp 400
Ile	Leu	Arg	Leu	Val 405	Ser	Asp	His	Arg	Asp 410	Ser	Gly	Val	Thr	Val 415	Asn
Gly	Glu	Leu	Ile 420	Gly	Ala	Pro		Pro 425	Pro	Asn	Gly		Lys 430	Lys	Gln

Arg Thr Tyr Leu Arg Thr Ile Thr Ile Leu Ile Asn Lys Pro Glu Arg 435 440 445 .

Ser Tyr Leu Glu Ile Thr Pro Ser Arg Val Ile Leu Asp Gly Gly Asp 450 455 460

Arg Leu Val Leu Pro Cys Asn Gln Ser Val Val Gly Ser Trp Gly 465 470 475 480

Leu Glu Val Ser Val Ser Ala Asn Ala Asn Val Thr Val Thr Ile Gln
485 490 495

Gly Ser Ile Ala Phe Val Ile Leu Ile His Leu Tyr Lys Lys Pro Ala 500 505 510

Pro Phe Gln Arg His His Leu Gly Phe Tyr Ile Ala Asn Ser Glu Gly 515 520 525

Leu Ser Ser Asn Cys His Gly Leu Leu Gly Gln Phe Leu Asn Gln Asp 530 540

Ala Arg Leu Thr Glu Asp Pro Ala Gly Pro Ser Gln Asn Leu Thr His 545 550 555 560

Pro Leu Leu Gln Val Gly Glu Gly Pro Glu Ala Val Leu Thr Val
565 570 575

Lys Gly His Gln Val Pro Val Val Trp Lys Gln Arg Lys Ile Tyr Asn 580 585 590

Gly Glu Glu Gln Xaa Asp Cys $\dot{\text{Trp}}$ Phe Ala Arg Asn Met Pro Pro Asn 595 600 605

<210> 211

<211> 252

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (252)

<223> Xaa equals stop translation

<400> 211

Met Ala Pro Ala Ser Arg Leu Leu Ala Leu Trp Ala Leu Ala Ala Val.

1 10 15

Ala Leu Pro Gly Ser Gly Ala Glu Gly Asp Gly Gly Trp Arg Pro Gly 20 25 30

Gly Pro Gly Ala Val Ala Glu Glu Glu Arg Cys Thr Val Glu Arg Arg

Ala Asp Leu Thr Tyr Ala Glu Phe Val Gln Gln Tyr Ala Phe Val Arg
50 60

Pro Val Ile Leu Gln Gly Leu Thr Asp Asn Ser Arg Phe Arg Ala Leu 65 70 75 80

Cys Ser Arg Asp Arg Leu Leu Ala Ser Phe Gly Asp Arg Val Val Arg

Leu Ser Thr Ala Asn Thr Tyr Ser Tyr His Lys Val Asp Leu Pro Phe 100 105 110

Gln Glu Tyr Val Glu Gln Leu Leu His Pro Gln Asp Pro Thr Ser Leu 115 120 125

Gly Asn Asp Thr Leu Tyr Phe Phe Gly Asp Asn Asn Phe Thr Glu Trp 130 140

Ala Ser Leu Phe Arg His Tyr Ser Pro Pro Pro Phe Gly Leu Leu Gly 145 150 155 160

Thr Ala Pro Ala Tyr Ser Phe Gly Ile Ala Gly Ala Gly Ser Gly Val

Pro Phe His Trp His Gly Pro Gly Tyr Ser Glu Val Ile Tyr Gly Arg 180 185 190

Lys Arg Trp Phe Leu Tyr Pro Pro Glu Lys Thr Pro Glu Phe His Pro 195 200 205

Asn Lys Thr Thr Leu Ala Trp Leu Arg Asp Thr Tyr Pro Ala Cys Thr 210 220

Val Cys Thr Ala Leu Glu Cys Thr Ile Arg Ala Gly Glu Val Leu Thr 225 230 235 240

Ser Arg Pro Leu Val Ala Cys Tyr Ala Gln Pro Xaa 245 250

<210> 212

<211> 226

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (226)

<223> Xaa equals stop translation

<400> 212

Met Lys Glu Ile Pro Ala Leu Leu His Leu Pro Val Leu Ile Ile Met

1 10 15

Ala Leu Ala Ile Leu Ser Phe Cys Tyr Gly Ala Gly Lys Ser Val His 20 25 30

Val Leu Arg His Ile Gly Gly Pro Glu Arg Glu Pro Pro Gln Ala Leu 35 40 45

Arg Pro Arg Asp Arg Arg Gln Glu Glu Ile Asp Tyr Arg Pro Asp

50 55 60

Gly Gly Ala Gly Asp Ala Asp Phe His Tyr Arg Gly Gln Met Gly Pro 80

Thr Glu Gln Gly Pro Tyr Ala Lys Thr Tyr Glu Gly Arg Arg Glu Ile 85 85 85 Leu Arg Phe Gln Thr Gly Asn Lys Ser 100

Pro Glu Val Leu Arg Ala Phe Asp Val Pro Asp Ala Glu Ala Arg Glu 115

His Pro Thr Val Val Pro Ser His Lys Ser Pro Val Leu Asp Thr Lys
130 135 140

Pro Lys Glu Thr Gly Gly Ile Leu Gly Glu Gly Thr Pro Lys Glu Ser 145 150 155 160

Ser Thr Glu Ser Ser Gln Ser Ala Lys Pro Val Ser Gly Gln Asp Thr 165 170 175

Ser Gly Asn Thr Glu Gly Ser Pro Ala Ala Glu Lys Ala Gln Leu Lys 180 185 190

Ser Glu Ala Ala Gly Ser Pro Asp Gln Gly Ser Thr Tyr Ser Pro Ala 195 200 205

Arg Gly Val Ala Gly Pro Arg Gly Gln Asp Pro Val Ser Ser Pro Cys 210 215 220

Gly Xaa 225

<210> 213 <211> 51

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (51)

<223> Xaa equals stop translation

<400> 213

Met Met Gly Leu Leu Glu Thr Gly Asn Val Leu Phe Trp Val Trp Val 1 5 10 15

Val Val Thr Cys Val Tyr Ser Leu Tyr Ala Asn Ser Leu Asn Cys Thr 20 25 30

Asp Met Asp Cys Ala Pro Phe Tyr Met Cys Val Met Leu Gln Gln Lys 35 40 45

Cys Gln Xaa 50

```
<210> 214
 <211> 172
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (172)
 <223> Xaa equals stop translation
 <400> 214
 Met Trp Leu Trp Ala Val Ser Pro Val Arg Pro Arg Thr Cys Leu Pro
 Pro Cys Pro Arg Leu Trp Leu Trp Ile Ser Met Thr Leu Val Pro Ser
                                 25
 Ser Ser Ala Trp Lys Ser His Gly Ala Pro Ser Thr Arg Met Thr Ser
                             40
 Pro Gln Leu Leu Leu Ser Thr Arg Pro Pro Gln Ser Pro Ser Ala
Ser Pro Pro Ile Ala Arg Ala His Arg Thr His Pro His Phe Gly Asn
Arg Leu Ser Ile Thr Cys Cys Asp Gly Arg Arg Ser Trp Arg Met Gly
                                     90
Gln His Gly Pro Cys His Leu Asn Leu Gln Thr Thr His Pro Ala His
Ser Ser Gln Ala Leu Pro Ala Thr His Gln Pro Leu Gly Pro Trp Cys
        115
Ser Ser Pro Ser Pro Phe Pro Ser Lys Leu Pro Ser Ala Gly Leu Arg
Pro Pro Ala Leu Gly Pro Trp Met Arg Arg Gly Pro Trp Pro Gln Ser
                    150
                                        155
Trp Gln Met Gly Met His Pro Thr Val Gly Leu Xaa
                165
<210> 215
<211> 48
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation
<400> 215
```

Met Trp Leu Leu Ile Ile Phe Cys Lys Ser Ala Ser Ala Ser Val Leu

```
Cys Trp Ile Lys Lys Phe His Pro Val Phe Gln Glu Ser Leu Leu Tyr 20 25 30
```

Leu Val Gln Glu Gly Ser Leu Cys Tyr Val Gln Gln Lys Val Pro Xaa 35 40 45

<210> 216

<211> 139

<212> PRT

<213> Homo sapiens

<400> 216

Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala Leu
1 5 10 15

Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu Glu Cys 20 25 30

Asp Tyr Ile Asn Ala Arg Ser Cys Cys Ser Lys Leu Asn Lys Trp Val 35 40 45

Ile Pro Glu Leu Ile Gly His Thr Ile Val Thr Val Leu Leu Met 50 55 60

Ser Leu His Trp Phe Ile Phe Leu Leu Asn Leu Pro Val Ala Thr Trp 65 70 75 80

Asn Ile Tyr Arg Tyr Ile Met Val Pro Ser Gly Asn Met Gly Val Phe 85 90 95

Asp Pro Thr Glu Ile His Asn Arg Gly Gln Leu Lys Ser His Met Lys
100 105 110

Glu Ala Met Ile Lys Leu Gly Phe His Leu Leu Cys Phe Phe Met Tyr 115 120 125

Leu Tyr Ser Met Ile Leu Ala Leu Ile Asn Asp 130 135

<210> 217

<211> 41

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (41)

<223> Xaa equals stop translation

<400> 217

Met Ser Gly Ser Ser Leu Pro Ser Ala Leu Ala Leu Ser Leu Leu 1 . 5

```
Val Ser Gly Ser Leu Leu Pro Gly Pro Gly Ala Ala Gln Asn Val Arg 20 25 30
```

Val Gln Ser Gly Gln Asp Gln Lys Xaa 35 40

<210> 218

<211> 52

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (52)

<223> Xaa equals stop translation

<400> 218

Met Pro Ser His Ile Arg Ala His Leu Phe Leu Leu Phe Phe Leu 1 5 10 15

Phe Ile Tyr Gln Gly Ile Ser Ser Ile Ser Gln Ala Ser Gly Leu Thr 20 25 30

Leu Lys Thr Gln Asn Glu Lys Asp Ile Gln Val Ser Ile Leu Lys Glu 35 40 45

Phe Val Val Xaa 50

<210> 219

<211> 49

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (49)

<223> Xaa equals stop translation

<400> 219

Met Cys Ile Tyr Gln Ser Glu Gln Met Leu Ala Leu Leu Leu Val Leu 1 5 10 15

Val Phe Cys Ile Ser Leu Leu Val Leu Val Cys Trp Gly Ser His Asn 20 25 30

Lys Val Pro Gln Lys Phe Ile Phe Ser Gln Phe Trp Gly Leu Glu Asp
35 40 45

Xaa

<210> 220

<211> 42

<212> PRT

<213> Homo sapiens

35

```
<220>
  <221> SITE
  <222> (42)
  <223> Xaa equals stop translation
  <400> 220
  Met Ala Val Pro Leu Phe Leu Tyr Ile Phe Thr Leu Leu Pro Leu Leu
                                       10
  Pro Phe Leu Leu Ser Leu Cys Phe Ser Pro Leu Thr Val Lys Arg Ser
 Ser Ser Ser Glu Ser Lys Ser Ser Leu Xaa
                               40
 <210> 221
 <211> 41
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (41)
 <223> Xaa equals stop translation
 Met Gly Met Leu Leu Ala Phe Trp Leu Pro Gly Ala Ser Trp Gln Glu
                                      10
Ala Gly Pro Arg Ala Ser Thr Gln Arg Met Arg Thr Gln Thr Gln Met
 Ser Thr Arg Lys Pro Lys Pro Ala Xaa
<210> 222
<211> 43
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (43)
<223> Xaa equals stop translation
<400> 222
Met Glu Pro Ala Met Val Leu Lys Phe Leu Ser Ser Leu Pro Glu Asn
                                     10
Leu Phe Leu Pro Ser Leu Leu Phe Phe Ala Trp Leu Cys Trp Asn Met
                                 25
Val Cys Gly Ser Pro Val Ser Cys Pro Tyr Xaa
```

<210> 223

<211> 204

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (204)

<223> Xaa equals stop translation

Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys

Phe Ser Gin Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys

Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val

Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala

Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile

Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser

Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala

Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr

Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu

Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly 155

Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala 165 170

Pro Gly Leu Ser Val Arg Leu Leu Arg Asp Pro Arg Cys Pro Asp Pro

Gly Cys Thr Ala Ala Pro Cys His Ala Ala His Xaa 195

<210> 224

<211> 43

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (43)

<223> Xaa equals stop translation

<400> 224

Met Arg Val Arg Ile Gly Leu Thr Leu Leu Leu Cys Ala Val Leu Leu 1 5 10 15

Ser Leu Ala Ser Ala Ser Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser 20 25 30

Leu Gly Phe Gln Asp Tyr Phe Asp Ile Arg Xaa 35 40

<210> 225

<211> 156

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (156)

<223> Xaa equals stop translation

<400> 225

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly

1 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly 20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys
35 40 45

Cys Met Asp Cys Ser Thr Ser Cys Pro Leu Pro Ala Ala Leu Ala His 50 55 60

Pro Trp Gly Arg Ser Glu Pro Asp Leu Arg Ala Gly Ala Ala Phe Trp 65 70 75 80

Leu Phe Gly Leu Glu Thr Met Pro Gln Glu Arg Glu Val His His Pro 95

His Arg Gly Asp Arg Arg Gly Leu Pro Ser Cys Gly Ala Asp Pro
100 105 110

Val Thr Met Cys Pro Leu Pro Ala Gly Ala Arg Pro Leu Ile Ile His 115 120 125

Ser Ser Ile Leu Glu Pro Val Ser Ala Ser Gln Thr Arg Arg Glu Pro 130 135 140

<210> 226

<211> 74

<212> PRT

<213> Homo sapiens

```
<220>
  <221> SITE
  <222> (38)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (48)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (54)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (55)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (68)
<223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (74)
<223> Xaa equals stop translation
 <400> 226
Met Phe Tyr Lys Leu Thr Leu Ile Leu Cys Glu Leu Ser Val Ala Gly
                                      10
Val Thr Gln Ala Ala Ser Gln Arg Pro Leu Gln Arg Leu Pro Arg His
             20
                                 25
Ile Cys Ser Gln Arg Xaa Pro Pro Gly Arg Cys Leu Leu Lys Ala Xaa
Leu Gln Thr Trp Xaa Xaa Pro Asp Lys Pro Ile Pro Arg Leu Ser
     50
                         55
Pro Pro Leu Xaa Ser Asp Pro Lys Arg Xaa
                     70
<210> 227
<211> 167
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (167)
<223> Xaa equals stop translation
<400> 227
```

Leu Ala Leu Pro Gly Ser Glu Ala Lys Asn Ser Gly Ala Ser Cys Pro 20 25 30

Pro Cys Pro Lys Tyr Ala Ser Cys His Asn Ser Thr His Cys Thr Cys 35 40 45

Glu Asp Gly Phe Arg Ala Arg Ser Gly Arg Thr Tyr Phe His Asp Ser 50 55 60

Ser Glu Lys Cys Glu Asp Ile Asn Glu Cys Glu Thr Gly Leu Ala Lys 65 70 75 80

Cys Lys Tyr Lys Ala Tyr Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys 85 90 . 95

Ser Cys Leu Val Lys Tyr Thr Leu Phe Asn Phe Leu Ala Gly Ile Ile 100 105 110

Asp Tyr Asp His Pro Asp Cys Tyr Glu Asn Asn Ser Gln Gly Thr Thr 115 120 125

Gln Ser Asn Val Asp Ile Trp Val Ser Gly Val Lys Pro Gly Phe Gly 130 135 140

Lys Gln Leu Val Arg Ile Thr Met Pro Phe Ser Tyr Pro Asn Ile Asn 150 155 160

Met Ser Ser Cys Asp Phe Xaa 165

<210> 228

<211> 71

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (71)

<223> Xaa equals stop translation

<400> 228

Met Lys Pro Lys His Leu Glu Trp Cys Leu Ala His Ser Trp Cys Val 1 5 10 15

Ile Trp Leu Ser Phe Val Ser Pro Pro Thr Ser His Leu Glu Cys Asp

Gly Phe Pro Gly Ser Leu Leu Pro Pro Cys Glu Glu Gly Arg Cys Phe 35 40 45

Pro Phe Thr Phe His His His Asp Cys His Gly Cys Ser Pro Leu Gln 50 55 60

Ser Ser Pro Gly Gln His Xaa 65 70

<210> 229 <211> 273 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (273) <223> Xaa equals stop translation <400> 229 Met Cys Cys Trp Pro Leu Leu Leu Leu Trp Gly Leu Leu Pro Gly Thr Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu Leu Asp Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly Ser Gln Ile Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val Gly Phe Gly Phe 55 Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile Val Val Gly Gly Val Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr Phe Thr Asn Ala Asn Arg 85 Glu Leu Lys Lys Asp Ala Gln Gln Asp Tyr His Leu Glu Tyr Ala Met Glu Asn Ser Thr His Thr Ile Ile Glu Phe Thr Arg Glu Leu His Thr 120 Cys Asp Ile Asn Asp Lys Ser Ile Thr Asp Ser Thr Val Arg Val Ile Trp Ala Tyr His His Glu Asp Ala Gly Glu Ala Gly Pro Lys Tyr His Asp Ser Asn Arg Gly Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys 165 170 Thr Ser Val Leu Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln Asp Val Pro Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe Lys Ile Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro 215 Val Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr

Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Pro Gly Ile Arg Ala Arg

245 250 255

Ile Ala Ile Thr Pro Thr Cys Pro Met His Ser Ser Pro Val Lys Leu 260 265 270

Xaa

<210> 230

<211> 82

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (82)

<223> Xaa equals stop translation

<400> 230

Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val Leu Leu 15

Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln Pro Val

Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr Phe His 35 40 45

Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala Cys Arg

Arg Gly Trp Arg Pro Ala Ser Gln His Arg Val Leu Lys Met Asn Arg 65 70 75

Asn Xaa

<210> 231

<211> 71

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (38)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 231

Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val 1 5 10 15

Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly
20 25 30

Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val 35 40 45

Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Asn Ile Trp Arg
50 55 60

Asp Leu Ala Cys Met Leu Ser 65 70

<210> 232

<211> 225

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (5)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 232

His Glu Arg Ala Xaa Gly Pro Ser Arg Gly His Gly Glu Leu Leu Ser 1 5 10

Cys Val Leu Gly Pro Arg Leu Tyr Lys Ile Tyr Arg Glu Arg Asp Ser 20 25 30

Glu Arg Ala Pro Ala Ser Val Pro Glu Thr Pro Thr Ala Val Thr Ala 35 40 45

Pro His Ser Ser Ser Trp Asp Thr Tyr Tyr Gln Pro Arg Ala Leu Glu 50 60

Lys His Ala Asp Ser Ile Leu Ala Leu Ala Ser Val Phe Trp Ser Ile 65 70 75 80

Ser Tyr Tyr Ser Ser Pro Phe Ala Phe Phe Tyr Leu Tyr Arg Lys Gly 85 90 95

Tyr Leu Ser Leu Ser Lys Val Val Pro Phe Ser His Tyr Ala Gly Thr 100 105 110

Leu Leu Leu Leu Ala Gly Val Ala Cys Ser Glu Ala Leu Ala Ala 115 120 125

Gly Pro Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser Trp Lys Gln His 130 135 140

Ile Gly Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu Pro Thr Thr Thr 145 150 155 160

Leu Thr Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly Lys Asn Pro Ala 165 170 175

Ala Gly Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly Val Trp Pro Cys 180 185 190

Phe Ala Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr Pro Ser Ser Thr 195 200 205

Gly Leu Arg Ser Cys Leu Val Arg Ser Pro Ala Thr Trp Trp Arg Thr 210 215 220

```
Pro
  225
 <210> 233
 <211> 314
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (147)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <220>
 <221> SITE
 <222> (211)
 <223> Xaa equals any of the naturally occurring L-amino acids
 <400> 233
 Met Leu Pro Ala Arg Leu Pro Phe Arg Leu Leu Ser Leu Phe Leu Arg
 Gly Ser Ala Pro Thr Ala Ala Arg His Gly Leu Arg Glu Pro Leu Leu
                                  25
Glu Arg Arg Cys Ala Ala Ala Ser Ser Phe Gln His Ser Ser Ser Leu
Gly Arg Glu Leu Pro Tyr Asp Pro Val Asp Thr Glu Gly Phe Gly Glu
                          55
Gly Gly Asp Met Gln Glu Arg Phe Leu Phe Pro Glu Tyr Ile Leu Asp
Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Gln
                                      90
Gln Gln Glu Glu Glu Arg Gln Arg Gln Gln Arg Arg Glu Glu Arg
                                 105
Arg Gln Gln Asn Leu Arg Ala Arg Ser Arg Glu His Pro Val Val Gly
                            120
His Pro Asp Pro Ala Leu Pro Pro Ser Gly Val Asn Cys Ser Gly Cys
Gly Ala Xaa Leu His Cys Gln Asp Ala Gly Val Pro Gly Tyr Leu Pro
                                        155
Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg
                165
Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu
                                185
Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala
                            200
```

Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu 210 215 220

Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 225 230 235 240

Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp 245 250 255

Ala Pro Gly Tyr Arg Gln Arg Leu Arg Glu Arg Leu Trp Glu Asp Cys 260 265 270

Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 275 280 285

Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Thr 290 295 300

Gly Pro Ala Gln Trp Ser Gly Thr Cys Gly 305

<210> 234

<211> 93

<212> PRT

<213> Homo sapiens

<400> 234

Met Arg Pro Gln Gly Pro Ala Ala Ser Pro Gln Arg Leu Arg Gly Leu 1 5 10 15

Leu Leu Leu Leu Leu Gln Leu Pro Ala Pro Ser Ser Ala Ser Glu 20 25 30

Ile Pro Lys Gly Lys Gln Lys Ala His Ser Gly Arg Gly Arg Trp Trp
35 40 45

Thr Cys Ile Met Glu Cys Ala Tyr Lys Gly Gln Gln Glu Cys Leu Val 50 60

Glu Thr Gly Ala Leu Gly Pro Met Ala Phe Arg Val His Leu Gly Ser
65 70 75 80

Gln Val Gly Met Asp Ser Lys Glu Lys Arg Gly Asn Val 85 90

<210> 235

<211> 73

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (73)

<223> Xaa equals stop translation

<400> 235

```
Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val
1 5 10 15 .
```

Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val 20 25 30

Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Thr Trp Lys 35 40 45

Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly Thr Cys Ser Ala
50 55 60

Lys Cys Thr Thr Arg Cys Trp Leu Xaa 65 70

<210> 236

<211> 349

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (283)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (293)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (325)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (326)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (349)

<223> Xaa equals stop translation

<400> 236

Met Leu Cys Pro Trp Arg Thr Ala Asn Leu Gly Leu Leu Leu Ile Leu 1 5 10 15

Thr Ile Phe Leu Val Ala Glu Ala Glu Gly Ala Ala Gln Pro Asn Asn 20 25 30

Ser Leu Met Leu Gln Thr Ser Lys Glu Asn His Ala Leu Ala Ser Ser 35 40 45

Ser Leu Cys Met Asp Glu Lys Gln Ile Thr Gln Asn Tyr Ser Lys Val

Leu 65	Ala	Glu	val	. Asn	Thr 70		Trp	Pro	Val	Lys 75		: Alá	a Thi	Asn	A1 8
Val	Leu	Cys	Cys	Pro 85	Pro	Ile	Ala	Leu	Arg 90		. Leu	ı Ile	e Ile	95	
Trp	Glu	Ile	11e		Arg	Gly	Gln	Pro 105		Cys	Thr	Lys	110		Ly
Lys	Glu	Thr 115		Glu	Thr	Lys	Glu 120		Asn	Cys	Thr	125		Arg	Ile
Thr	Trp 130	Val	Ser	Arg	Pro	Asp 135		Asn	Ser	Asp	Leu 140		Ile	Arg	Thi
Val 145	Ala	Ile	Thr	His	Asp 150	Gly	Tyr	Tyr	Arg	Cys 155		Met	Val	Thr	Pro 160
Asp	Gly	Asn	Phe	His 165	Arg	Gly	Tyr	His	Leu 170		Val	Leu	'Val	Thr 175	Pro
Glu	Val	Thr	Leu 180	Phe	Gln	Asn	Arg	Asn 185	Arg	Thr	Ala	Val	Cys 190	Lys	Ala
Val	Ala	Gly 195	Lys	Pro	Ala	Ala	His 200	Ile	Ser	Trp	Ile	Pro 205	Glu	Gly	Asp
Cys	Ala 210	Thr	Lys	Gln	Glu	Туг 215	Trp	Ser	Asn	Gly	Thr 220	Val	Thr	Val	Lys
Ser 225	Thr	Cys	His	Trp	Glu 230	Val	His	Asn	Val	Ser 235	Thr	Val	Asn	Суѕ	His 240
/al	Ser	His	Leu	Thr 245	Gly	Asn	Lys	Ser	Leu 250	Tyr	Ile	Glu	Leu	Leu 255	Pro
/al	Pro	Gly	Ala 260	Lys	Lys	Ser	Ser	Lys 265	Leu	Tyr	Ile	Pro	Туг 270	Ile	Ile
Leu	Thr	Ile 275	Ile	Ile	Leu	Thr	Ile 280	Val	Gly	Xaa	Ile	Trp 285	Leu	Leu-	Lys
/al	Asn 290	Gly	Cys	Xaa	Lys	Tyr 295	Lys	Leu	Asn	Lys	Pro 300	Glu	Ser	Thr	Pro
7al 805	Val	Glu	Glu	Asp	Glu 310	Met	Gln	Pro	Tyr	Ala 315	Phe	Tyr	Thr	Glu	Lys 320
sn	Asn	Pro	Leu	Xaa 325	Xaa	Thr	Thr	Asn	Lys 330	Val	Lys	Ala	Ser	Glu 335	Ala

Leu Gln Ser Glu Val Asp Thr Asp Leu His Thr Leu Xaa

345

<210> 237 <211> 17

340

<212> PRT <213> Homo sapiens

```
Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr Ser Gly Trp Asp Thr Leu
                                     10
 Asn
 <210> 238
 <211> 14
 <212> PRT
 <213> Homo sapiens
 <400> 238
 Val Thr Glu Glu Ile Lys Asn Pro Glu Arg Asn Leu Pro Leu
 1
                       . 10
                  5
<210> 239
<211> 9
<212> PRT
<213> Homo sapiens
<400> 239
Ile Gly Ile Ser Met Pro Ile Val Thr
                 5
<210> 240
<211> 13
<212> PRT
<213> Homo sapiens
<400> 240
Ile Tyr Ile Leu Thr Asn Val Ala Tyr Tyr Thr Val Leu
               5
<210> 241
<211> 11
<212> PRT
<213> Homo sapiens
<400> 241
Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln
                                    10
<210> 242
<211> 13
<212> PRT
<213> Homo sapiens
Val Ala Leu Ser Cys Phe Gly Gly Leu Asn Ala Ser Ile
 1 .
                5
```

```
<210> 243
  <211> 15
  <212> PRT
  <213> Homo sapiens
  <400> 243
  Ser Arg Leu Phe Phe Val Gly Ser Arg Glu Gly His Leu Pro Asp
                   5
                                       10
 <210> 244
 <211> 11
 <212> PRT
 <213> Homo sapiens
 <400> 244
 Ser Phe Ser Tyr Trp Phe Phe Val Gly Leu Ser
 <210> 245
 <211> 11
 <212> PRT
<213> Homo sapiens
 <400> 245
 Val Gly Gln Leu Tyr Leu Arg Trp Lys Glu Pro
 <210> 246
 <211> 16
 <212> PRT
<213> Homo sapiens
 <400> 246
 Arg Pro Arg Pro Leu Lys Leu Ser Val Phe Pro Ile Val Phe Cys
                                      10
 <210> 247
 <211> 9
 <212> PRT
 <213> Homo sapiens
 <400> 247
 Asp Thr Ile Asn Ser Leu Ile Gly Ile
 <210> 248
 <211> 44
 <212> PRT
 <213> Homo sapiens
 <400> 248
```

- Ala Thr Ala Leu Pro Pro Lys Ile Val Gly Ser Ala Thr Arg Tyr Leu 1 5 10 15
- Gln Val Leu Cys Met Ser Val Ala Ala Glu Met Asp Leu Glu Asp Gly
 20 25 30
- Gly Glu Met Pro Lys Gln Arg Asp Pro Lys Ser Asn 35
- <210> 249
- <211> 352
- <212> PRT
- <213> Homo sapiens
- <400> 249
- Leu Leu Ala Ala Ala Cys Ile Cys Leu Leu Thr Phe Ile Asn Cys Ala 1 5 10
- Tyr Val Lys Trp Gly Thr Leu Val Gln Asp Ile Phe Thr Tyr Ala Lys
 20 25 30
- Val Leu Ala Leu Ile Ala Val Ile Val Ala Gly Ile Val Arg Leu Gly
 35 40
- Gln Gly Ala Ser Thr His Phe Glu Asn Ser Phe Glu Gly Ser Ser Phe
 50 60
- Ala Val Gly Asp Ile Ala Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr 65 70 75 80
- Ser Gly Trp Asp Thr Leu Asn Tyr Val Thr Glu Glu Ile Lys Asn Pro
 85 90 95
- Glu Arg Asn Leu Pro Leu Ser Ile Gly Ile Ser Met Pro Ile Val Thr 100 105 110
- Ile Ile Tyr Ile Leu Thr Asn Val Ala Tyr Tyr Thr Val Leu Asp Met 115 120 125
- Arg Asp Ile Leu Ala Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln 130 135 140
- Ile Phe Gly Ile Phe Asn Trp Ile Ile Pro Leu Ser Val Ala Leu Ser 145 150 155 160
- Cys Phe Gly Gly Leu Asn Ala Ser Ile Val Ala Ala Ser Arg Leu Phe 165 170 175
- Phe Val Gly Ser Arg Glu Gly His Leu Pro Asp Ala Ile Cys Met Ile 180 185 190
- His Val Glu Arg Phe Thr Pro Val Pro Ser Leu Leu Phe Asn Gly Ile 195 200 205
- Met Ala Leu Ile Tyr Leu Cys Val Glu Asp Ile Phe Gln Leu Ile Asn 210 215 220
- Tyr Tyr Ser Phe Ser Tyr Trp Phe Phe Val Gly Leu Ser Ile Val Gly

173

225 230 235 240

Gln Leu Tyr Leu Arg Trp Lys Glu Pro Asp Arg Pro Arg Pro Leu Lys 245 250 250

Leu Ser Val Phe Phe Pro Ile Val Phe Cys Leu Cys Thr Ile Phe Leu 260 265 270

Val Ala Val Pro Leu Tyr Ser Asp Thr Ile Asn Ser Leu Ile Gly Ile 275 280 285

Ala Ile Ala Leu Ser Gly Leu Pro Phe Tyr Phe Leu Ile Ile Arg Val 290 295 300

Pro Glu His Lys Arg Pro Leu Tyr Leu Arg Arg Ile Val Gly Ser Ala 305 310 315 320

Thr Arg Tyr Leu Gln Val Leu Cys Met Ser Val Ala Ala Glu Met Asp 325 330 335

Leu Glu Asp Gly Glu Met Pro Lys Gln Arg Asp Pro Lys Ser Asn 340 345 350

<210> 250

<211> 119

<212> PRT

<213> Homo sapiens

<400> 250

Ala Ala Arg Gly Ser Gly Val Arg Asp Pro Leu Glu Glu Ala Val Cys

1 10 15

Pro Phe Ser Asp Leu Gln Leu His Ala Gly Arg Thr Thr Ala Leu Phe 20 25 30

Lys Ala Val Arg Gln Gly His Leu Ser Leu Gln Arg Leu Leu Ser
35 40 45

Phe Val Cys Leu Cys Pro Ala Pro Arg Gly Gly Ala Tyr Arg Gly Arg 50 60

Gln Ala Ser Leu Ser Cys Gly Gly Leu His Pro Val Arg Ala Ser Arg 65 70 75 80

Leu Leu Cys Leu Pro Lys Gln Ala Trp Ala Met Ala Gly Ala Pro Pro
85 90 95

Pro Val Ser Leu Pro Pro Cys Ser Leu Ile Ser Asp Cys Cys Ala Ser 100 105 110

Asn Gln Arg Asp Ser Val Gly 115

<210> 251

<211> 356 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (37) <223> Xaa equals any of the naturally occurring L-amino acids <220> <221> SITE <222> (280) <223> Xaa equals any of the naturally occurring L-amino acids <400> 251 Leu Ser Lys Ala Phe Leu Asp Ser Pro Asn Arg Leu Leu Ala Val Glu 5 Met Asn Thr Asp His Leu Arg Leu Thr Val Pro Asn Gly Ile Gly Ala 25 Leu Lys Leu Arg Xaa Met Glu His Tyr Phe Ser Gln Gly Leu Ser Val Gln Leu Phe Asn Asp Gly Ser Lys Gly Lys Leu Asn His Leu Cys Gly Ala Asp Phe Val Lys Ser His Gln Lys Pro Pro Gln Gly Met Glu Ile Lys Ser Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile 90 Val Tyr Tyr His Asp Ala Asp Gly His Phe His Leu Ile Asp Gly 105 Asp Lys Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu Leu Val 120 Glu Ile Gly Glu Ser Leu Asn Ile Gly Val Val Gln Thr Ala Tyr Ala 130 Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu Val Met Lys Val Pro Val 150 155 Tyr Cys Thr Lys Thr Gly Val Lys His Leu His His Lys Ala Gln Glu Phe Asp Ile Gly Val Tyr Phe Glu Ala Asn Gly His Gly Thr Ala Leu 185 Phe Ser Thr Ala Val Glu Met Lys Ile Lys Gln Ser Ala Glu Gln Leu 200 Glu Asp Lys Lys Arg Lys Ala Ala Lys Met Leu Glu Asn Ile Ile Asp 220 Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile Ser Asp Met Leu Val Ile

175

Glu Ala Ile Leu Ala Leu Lys Gly Leu Thr Val Gln Gln Trp Asp Ala

Leu Tyr Thr Asp Leu Pro Asn Arg Gln Leu Lys Val Gln Val Ala Asp

Arg Arg Val Ile Ser Thr Thr Xaa Ala Glu Arg Gln Ala Val Thr Pro 280 285

Pro Gly Leu Gln Glu Ala Ile Asn Asp Leu Val Lys Lys Tyr Lys Leu 295

Ser Arg Ala Phe Val Arg Pro Ser Gly Thr Glu Asp Val Val Arg Val

Tyr Ala Glu Ala Asp Ser Gln Glu Ser Ala Asp His Leu Ala His Glu 330

Val Ser Leu Ala Val Phe Gln Leu Ala Gly Gly Ile Gly Glu Arg Pro

Gln Pro Gly Phe 355

<210> 252

<211> 26

<212> PRT

<213> Homo sapiens

<400> 252

Leu Ser Lys Ala Phe Leu Asp Ser Pro Asn Arg Leu Leu Ala Val Glu

Met Asn Thr Asp His Leu Arg Leu Thr Val 20

<210> 253

<211> 28

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (11)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 253

Pro Asn Gly Ile Gly Ala Leu Lys Leu Arg Xaa Met Glu His Tyr Phe 1 10

Ser Gln Gly Leu Ser Val Gln Leu Phe Asn Asp Gly

<210> 254

<211> 28

```
<212> PRT
   <213> Homo sapiens
   <400> 254
   Ser Lys Gly Lys Leu Asn His Leu Cys Gly Ala Asp Phe Val Lys Ser
  His Gln Lys Pro Pro Gln Gly Met Glu Ile Lys Ser
  <210> 255
  <211> 28
  <212> PRT
  <213> Homo sapiens
  <400> 255
  Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile Val Tyr
  Tyr Tyr His Asp Ala Asp Gly His Phe His Leu Ile
 <210> 256
 <211> 28
 <212> PRT
 <213> Homo sapiens
 <400> 256
 Asp Gly Asp Lys Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu
 Leu Val Glu Ile Gly Glu Ser Leu Asn Ile Gly Val
              20
 <210> 257
 <211> 28
 <212> PRT
 <213> Homo sapiens
<400> 257
Val Gln Thr Ala Tyr Ala Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu
                                      10
Val Met Lys Val Pro Val Tyr Cys Thr Lys Thr Gly
                                 25
<210> 258
<211> 28
<212> PRT
<213> Homo sapiens
<400> 258
Val Lys His Leu His His Lys Ala Gln Glu Phe Asp Ile Gly Val Tyr
```

```
Phe Glu Ala Asn Gly His Gly Thr Ala Leu Phe Ser
              20
                                  25
 <210> 259
 <211> 28
 <212> PRT
 <213> Homo sapiens
<400> 259
Thr Ala Val Glu Met Lys Ile Lys Gln Ser Ala Glu Gln Leu Glu Asp
                                      10
Lys Lys Arg Lys Ala Ala Lys Met Leu. Glu Asn Ile
             20
<210> 260
<211> 28
<212> PRT
<213> Homo sapiens
<400> 260
Ile Asp Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile Ser Asp Met Leu
Val Ile Glu Ala Ile Leu Ala Leu Lys Gly Leu Thr
             20
                                  25
<210> 261
<211> 28
<212> PRT
<213> Homo sapiens
<400> 261
Val Gln Gln Trp Asp Ala Leu Tyr Thr Asp Leu Pro Asn Arg Gln Leu
Lys Val Gln Val Ala Asp Arg Arg Val Ile Ser Thr
             20
                                  25
<210> 262
<211> 28
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (2)
<223> Xaa equals any of the naturally occurring L-amino acids
<400> 262
Thr Xaa Ala Glu Arg Gln Ala Val Thr Pro Pro Gly Leu Gln Glu Ala
                                     10
Ile Asn Asp Leu Val Lys Lys Tyr Lys Leu Ser Arg
            20
                                 25
```

```
<210> 263
```

<211> 24

<212> PRT

<213> Homo sapiens

<400> 263

Ala Phe Val Arg Pro Ser Gly Thr Glu Asp Val Val Arg Val Tyr Ala

1 5 10 15

Glu Ala Asp Ser Gln Glu Ser Ala 20

<210> 264

<211> 26

<212> PRT

<213> Homo sapiens

<400> 264

Asp His Leu Ala His Glu Val Ser Leu Ala Val Phe Gln Leu Ala Gly

1 5 10 15

Gly Ile Gly Glu Arg Pro Gln Pro Gly Phe

<210> 265

<211> 443

<212> PRT

<213> Homo sapiens

<400> 265

Gly Thr Arg Ala Ala Pro Gly Leu Gly Ala Trp Gly Arg Arg Ser Pro

1 5 10 15

Pro Ser Phe Ser Pro Pro Arg Pro Arg Arg Pro Gly Val Met Ala Gly 20 25 30

Leu Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu Leu Leu Gly 35 40 45

Ala Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile Ala Leu Pro 50 55 60

Arg Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr Pro Thr Leu 65 70 75 80

Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser Lys Arg His Ile Thr Met 85 90 95

Leu Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe Ser Cys Gln
100 105 110

Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn Ile Asp Cys 115 120 125

Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln Pro Ser Thr

	130)				135	5				140)			
Se:		ı Lei	u Pro	Th	r Leu 150		n Arg	Thr	: Phe	11e) As	o Va	l Ly:	s Ala 160
His	s Lys	S Sei	r Ile	Gly 165	/ Leu	Glu	ı Leu	Glr	170		Ile	e Pro	o Ar	g Lei 179	_
Glr	ı Ile	e Gly	/ Pro 180		/ Glu	Ser	Cys	Pro 185		Gly	Val	. Thi	His 190		: Ile
Ser	Gly	7 Arg 195		e Asp	Ala	Thr	Val 200		Arg	Ile	Gly	7 Thr 205		e Cys	Sei
Asn	210		· Val	Ser	Arg	Ile 215		Met	Gln	Glu	Gly 220		. Lys	Met	Ala
Leu 225		Leu	Pro	Trp	Phe 230	His	Pro	Arg	Asn	Val 235		Gly	Phe	e Ser	240
Ala	Asn	Arg	Ser	Ser 245	Ile	Lys	Arg	Leu	Cys 250		Ile	Glu	Ser	Val 255	
Glu	Gly	Glu	Gly 260		Ala	Thr	Leu	Met 265	Ser	Ala	Asn	Tyr	Pro 270		Gly
Phe	Pro	Glu 275		Glu	Leu	Met	Thr 280	Trp	Gln	Phe	Val	Val 285		Ala	His
Leu	Arg 290		Ser	Val	Ser	Phe 295	Leu	Asn	Phe	Asn	Leu 300	Ser	Asn	Суз	Glu
Arg 305		Glu	Glu	Arg	Val 310	Glu	Tyr	Tyr	Ile	Pro 315	Gly	Ser	Thr	Thr	Asn 320
Pro	Glu	Val	Phe	Lys 325	Leu	Glu	Asp	Lys	Gln 330	Pro	Gly	Asn	Met	Ala 335	Gly
Asn	Phe	Asn	Leu 340	Ser	Leu	Gln	Gly	Cys 345	Asp	Gln	Asp	Ala	Gln 350	Ser	Pro
Gly	Ile	Leu 355	Arg	Leu	Gln	Phe	Gln 360	Val	Leu	Val	Gln	His 365	Pro	Gln	Asn
Glu	Ser 370	Asn	Lys	Ile	Tyr	Val 375	Val	Asp	Leu	Ser	Asn 380	Glu	Arg	Ala	Met
Ser 385	Leu	Thr	Ile	Glu	Pro 390	Arg	Pro	Val	Lys	Gln 395	Ser	Arg	Lys	Phe	Val 400
Pro	Gly	Cys	Phe	Val 405	Cys	Leu	Glu	Ser	Arg 410	Thr	Cys	Ser	Ser	Asn 415	Leu
Thr	Leu	Thr	Ser 420	Gly	Ser	Lys		Lys 425	Ile	Ser	Phe	Leu	Cys 430	Asp	Asp
·Leu	Thr	Arg 435	Leu	Trp	Met .		Val 440	Glu	Lys	Pro					

```
<210> 266
<211> 159
```

<212> PRT

<213> Homo sapiens

<400> 266

Phe Glu Ile Ala Leu Pro Arg Glu Ser Asn Ile Thr Val Leu Ile Lys

1 10 15

Leu Gly Thr Pro Thr Leu Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser

Lys Arg His Ile Thr Met Leu Ser Ile Lys Ser Gly Glu Arg Ile Val 35 40 45

Phe Thr Phe Ser Cys Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile 50 60

Gln Lys Asn Ile Asp Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val
65 70 75 80

Gln Leu Gln Pro Ser Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe
85 90 95

Ile Trp Asp Val Lys Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe
100 105 110

Ser Ile Pro Arg Leu Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp 115 120 125

Gly Val Thr His Ser Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg
130 135 140

Ile Gly Thr Phe Cys Ser Asn Gly Thr Val Ser Arg Ile Lys Met
145 150 155

<210> 267

<211> 9

<212> PRT

<213> Homo sapiens

<400> 267

Phe Val Arg Asp Pro Phe Val Arg Leu

<210> 268

<211> 13

<212> PRT

<213> Homo sapiens

<400> 268

Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser

<210> 269

<211> 15

<212> PRT

<213> Homo sapiens

<400> 269

Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe

<210> 270

<211> 380

<212> PRT

<213> Homo sapiens

<400> 270

Tyr Leu His Thr Ser Phe Ser Arg Pro His Thr Gly Pro Pro Leu Pro

Thr Pro Gly Pro Asp Arg Asp Arg Glu Leu Thr Ala Asp Ser Asp Val 25

Asp Glu Phe Leu Asp Lys Phe Leu Ser Ala Gly Val Lys Gln Ser Asp

Leu Pro Arg Lys Glu Thr Glu Gln Pro Pro Ala Pro Gly Ser Met Glu

Glu Asn Val Arg Gly Tyr Asp Trp Ser Pro Arg Asp Ala Arg Arg Ser

Pro Asp Gln Gly Arg Gln Gln Ala Glu Arg Arg Ser Val Leu Arg Gly

Phe Cys Ala Asn Ser Ser Leu Ala Phe Pro Thr Lys Glu Arg Ala Phe 105

Asp Asp Ile Pro Asn Ser Glu Leu Ser His Leu Ile Val Asp Asp Arg

His Gly Ala Ile Tyr Cys Tyr Val Pro Lys Val Ala Cys Thr Asn Trp 130 135

Lys Arg Val Met Ile Val Leu Ser Gly Ser Leu Leu His Arg Gly Ala 150

Pro Tyr Arg Asp Pro Leu Arg Ile Pro Arg Glu His Val His Asn Ala

Ser Ala His Leu Thr Phe Asn Lys Phe Trp Arg Arg Tyr Gly Lys Leu 185

Ser Arg His Leu Met Lys Val Lys Leu Lys Lys Tyr Thr Lys Phe Leu 200

Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe Arg Ser Lys

Phe Glu Leu Glu Asn Glu Glu Phe Tyr Arg Lys Phe Ala Val Pro Met 235

- Ala Phe Arg Ala Gly Leu Lys Val Ser Phe Ala Asn Phe Ile Gln Tyr 260 265 270
- Leu Leu Asp Pro His Thr Glu Lys Leu Ala Pro Phe Asn Glu His Trp 275 280 285
- Arg Gln Val Tyr Arg Leu Cys His Pro Cys Gln Ile Asp Tyr Asp Phe 290 295 300
- Val Gly Lys Leu Glu Thr Leu Asp Glu Asp Ala Ala Gln Leu Leu Gln 305 310 315 320
- Leu Leu Gln Val Asp Arg Gln Leu Arg Phe Pro Pro Ser Tyr Arg Asn 325 330 335
- Arg Thr Ala Ser Ser Trp Glu Glu Asp Trp Phe Ala Lys Ile Pro Leu 340 345 350
- Ala Trp Arg Gln Gln Leu Tyr Lys Leu Tyr Glu Ala Asp Phe Val Leu 355 360 365
- Phe Gly Tyr Pro Lys Pro Glu Asn Leu Leu Arg Asp 370 380
- <210> 271
- <211> 274
- <212> PRT
- <213> Homo sapiens
- <400> 271
- Lys Leu Val Arg Leu Gln Val Pro Val Arg Asn Ser Arg Val Asp Pro

 1 5 10 15
- Arg Val Arg Ser Lys Ile Gly Ser Arg Arg Trp Met Leu Gln Leu Ile 20 25 30
- Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys 35 40 45
- Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys
 50 55 60
- Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val 65 70 75 80
- Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala 85 90 95
- Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile 100 105 110
- Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser 115 120 125

183 Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala 135 Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr 150 155 Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly 185 Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala 200 Leu Ala Phe Leu Ser Gly Tyr Tyr Val Thr Leu Ala Ala Gln Ile Leu Ala Val Leu Leu Pro Pro Val Met Leu Leu Ile Asp Gly Asn Val Ala 235 Tyr Trp His Asn Thr Arg Arg Val Glu Phe Trp Asn Gln Met Lys Leu Leu Gly Glu Ser Val Gly Ile Phe Gly Thr Ala Val Ile Leu Ala Thr 265 Asp Gly <210> 272

<211> 203

<212> PRT

<213> Homo sapiens

<400> 272

Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys 10

Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys

Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val

Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala

Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile

Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser 85

Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala 105

Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr

115 120 125 Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu 135 Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly 155 Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala 165 170 Pro Gly Leu Ser Val Arg Leu Leu Arg Asp Pro Arg Cys Pro Asp Pro 185 Gly Cys Thr Ala Ala Pro Cys His Ala Ala His 195 200 <210> 273 <211> 407 <212> PRT <213> Homo sapiens <400> 273 Ser Asn Glu Ile Leu Leu Ser Phe Pro Gln Asn Tyr Tyr Ile Gln Trp Leu Asn Gly Ser Leu Ile His Gly Leu Trp Asn Leu Ala Ser Leu Phe Ser Asn Leu Cys Leu Phe Val Leu Met Pro Phe Ala Phe Phe Leu Glu Ser Glu Gly Phe Ala Gly Leu Lys Lys Gly Ile Arg Ala Arg Ile 55 Leu Glu Thr Leu Val Met Leu Leu Leu Leu Ala Leu Leu Ile Leu Gly 70 Ile Val Trp Val Ala Ser Ala Leu Ile Asp Asn Asp Ala Ala Ser Met 90 Glu Ser Leu Tyr Asp Leu Trp Glu Phe Tyr Leu Pro Tyr Leu Tyr Ser 100 Cys Ile Ser Leu Met Gly Cys Leu Leu Leu Leu Cys Thr Pro Val 120 Gly Leu Ser Arg Met Phe Thr Val Met Gly His Leu Leu Val Lys Pro 135 Thr Ile Leu Glu Asp Leu Asp Glu Gln Ile Tyr Ile Ile Thr Leu Glu 150 155 Glu Glu Ala Leu Gln Arg Arg Leu Asn Gly Leu Ser Ser Ser Val Glu 170

Tyr Asn Ile Met Glu Leu Glu Gln Glu Leu Glu Asn Val Lys Thr Leu

185

180

WO 00/06698 PCT/US99/17130

185

Lys Thr Lys Leu Glu Arg Arg Lys Lys Ala Ser Ala Trp Glu Arg Asn 195 200 205

Leu Val Tyr Pro Ala Val Met Val Leu Leu Leu Ile Glu Thr Ser Ile 210 215 220

Ser Val Leu Leu Val Ala Cys Asn Ile Leu Cys Leu Leu Val Asp Glu 225 230 235 240

Thr Ala Met Pro Lys Gly Thr Arg Gly Pro Gly Ile Gly Asn Ala Ser 245 250 255

Leu Ser Thr Phe Gly Phe Val Gly Ala Ala Leu Glu Ile Ile Leu Ile 260 265 270

Phe Tyr Leu Met Val Ser Ser Val Val Gly Phe Tyr Ser Leu Arg Phe 275 280 285

Phe Gly Asn Phe Thr Pro Lys Lys Asp Asp Thr Thr Met Thr Lys Ile 290 295 300

Ile Gly Asn Cys Val Ser Ile Leu Val Leu Ser Ser Ala Leu Pro Val 305 310 315 320

Met Ser Arg Thr Leu Gly Ile Thr Arg Phe Asp Leu Leu Gly Asp Phe 325 330 335

Gly Arg Phe Asn Trp Leu Gly Asn Phe Tyr Ile Val Leu Ser Tyr Asn 340 345 350

Leu Leu Phe Ala Ile Val Thr Thr Leu Cys Leu Val Arg Lys Phe Thr 355 360 365

Ser Ala Val Arg Glu Glu Leu Phe Lys Ala Leu Gly Leu His Lys Leu 370 380

His Leu Pro Asn Thr Ser Arg Asp Ser Glu Thr Ala Lys Pro Ser Val 385 390 395 400

Asn Gly His Gln Lys Ala Leu 405

<210> 274

<211> 165

<212> PRT

<213> Homo sapiens

<400> 274

Arg Ser Tyr Met Gln Ser Val Trp Thr Glu Glu Ser Gln Cys Thr Leu

1 10 15

Leu Asn Ala Ser Ile Thr Glu Thr Phe Asn Cys Ser Phe Ser Cys Gly 20 25 30

Pro Asp Cys Trp Lys Leu Ser Gln Tyr Pro Cys Leu Gln Val Tyr Val 35 40

- Asn Leu Thr Ser Ser Gly Glu Lys Leu Leu Leu Tyr His Thr Glu Glu 50 55 60
- Thr Ile Lys Ile Asn Gln Lys Cys Ser Tyr Ile Pro Lys Cys Gly Lys 65 70. 75 80
- Asn Phe Glu Glu Ser Met Ser Leu Val Asn Val Val Met Glu Asn Phe
 85 90 95
- Arg Lys Tyr Gln His Phe Ser Cys Tyr Ser Asp Pro Glu Gly Asn Gln 100 105 110
- Lys Ser Val Ile Leu Thr Lys Leu Tyr Ser Ser Asn Val Leu Phe His 115 120 125
- Ser Leu Phe Trp Pro Thr Cys Met Met Ala Gly Gly Val Ala Ile Val 130 135 140
- Ala Met Val Lys Leu Thr Gln Tyr Leu Ser Leu Leu Cys Glu Arg Ile 145 - 150 - 155 - 160
- Gln Arg Ile Asn Arg 165
- <210> 275
- <211> 155
- <212> PRT
- <213> Homo sapiens
- <400> 275
- Ala Phe Ala His Leu Gln Leu Gly Pro Met Trp Lys Leu Trp Arg Ala

 1 5 10 15
- Glu Glu Gly Ala Ala Ala Leu Gly Gly Ala Leu Phe Leu Leu Phe 20 25 30
- Ala Leu Gly Val Arg Gln Leu Leu Lys Gln Arg Arg Pro Met Gly Phe 35 40 45
- Pro Pro Gly Pro Pro Gly Leu Pro Phe Ile Gly Asn Ile Tyr Ser Leu 50 55 60
- Ala Ala Ser Ser Glu Leu Pro His Val Tyr Met Arg Lys Gln Ser Gln 65 70 75 80
- Val Tyr Gly Glu Val Gln Pro Arg Arg Ala Pro Gly Arg Glu Gly Arg 85 90 95
- Gln Ala Gly Pro Gly Trp Pro Gly Pro Ser Trp Leu Asp Leu Trp Pro 100 105 110
- Pro Leu Gly Arg Leu Val Gly Thr Ser Pro Cys Ala Gly Cys Pro Leu 115 120 125
- Arg Asp Thr Arg Phe Pro Gly Leu Glu Gly Arg Ser Pro Arg Arg Arg 130 135 140
- Ala Pro Leu Gln Gly Glu Pro Arg Pro Cys Arg

187

145 150 155

<210> 276

<211> 42

<212> PRT

<213> Homo sapiens

<400> 276

Met Arg Val Arg Ile Gly Leu Thr Leu Leu Cys Ala Val Leu Leu 1 5 10 15

Ser Leu Ala Ser Ala Ser Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser 20 25. 30

Leu Gly Phe Gln Asp Tyr Phe Asp Ile Arg

<210> 277

<211> 155

<212> PRT

<213> Homo sapiens

<400> 277

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly
1 5 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly 20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys 35 40 45

Cys Met Asp Cys Ser Thr Ser Cys Pro Leu Pro Ala Ala Leu Ala His 50 55 60

Pro Trp Gly Arg Ser Glu Pro Asp Leu Arg Ala Gly Ala Ala Phe Trp 65 70 75 80

Leu Phe Gly Leu Glu Thr Met Pro Gln Glu Arg Glu Val His His Pro 85 90 95

His Arg Gly Asp Arg Arg Gly Leu Pro Ser Cys Gly Ala Asp Pro 100 105 110

Val Thr Met Cys Pro Leu Pro Ala Gly Ala Arg Pro Leu Ile His 115 120 125

Ser Ser Ile Leu Glu Pro Val Ser Ala Ser Gln Thr Arg Arg Glu Pro 130 135 140

Ser Ser Ser Asn His Lys Gly Gly Gly Arg 145 150 155

<210> 278

<211> 207

<212> PRT

<213> Homo sapiens

<400> 278

Gly Thr Ser Phe Leu Asp Pro Thr Leu Ser Leu Phe Val Leu Glu Lys

1 5 10 15

Phe Asn Leu Pro Ala Gly Tyr Val Gly Leu Val Phe Leu Gly Met Ala 20 25 30

Leu Ser Tyr Ala Ile Ser Ser Pro Leu Phe Gly Leu Leu Ser Asp Lys
35 40 45

Arg Pro Pro Leu Arg Lys Trp Leu Leu Val Phe Gly Asn Leu Ile Thr 50 55 60

Ala Gly Cys Tyr Met Leu Leu Gly Pro Val Pro Ile Leu His Ile Lys
65 70 75 80

Ser Gln Leu Trp Leu Leu Val Leu Ile Leu Val Val Ser Gly Leu Ser 85 90 95

Ala Gly Met Ser Ile Ile Pro Thr Phe Pro Glu Ile Leu Ser Cys Ala 100 105 110

His Glu Asn Gly Phe Glu Glu Gly Leu Ser Thr Leu Gly Leu Val Ser 115 120 125

Gly Leu Phe Ser Ala Met Trp Ser Ile Gly Ala Phe Met Gly Pro Thr 130 140

Leu Gly Gly Phe Leu Tyr Glu Lys Ile Gly Phe Glu Trp Ala Ala Ala 145 150 155 160

Ile Gln Gly Leu Trp Ala Leu Ile Ser Gly Leu Ala Met Gly Leu Phe
165 170 175

Tyr Leu Leu Glu Tyr Ser Arg Arg Lys Arg Ser Lys Ser Gln Asn Ile 180 185 190

Leu Ser Thr Glu Glu Glu Arg Thr Thr Leu Leu Pro Asn Glu Thr

<210> 279

<211> 85

<212> PRT

<213> Homo sapiens

<400> 279

Gly Thr Arg Glu Ala Arg Leu Arg Asp Leu Thr Arg Phe Tyr Asp Lys

1 10 15

Val Leu Ser Leu His Glu Asp Ser Thr Thr Pro Val Ala Asn Pro Leu 20 25 30

Leu Ala Phe Thr Leu Ile Lys Arg Leu Gln Ser Asp Trp Arg Asn Val

Val His Ser Leu Glu Ala Ser Glu Asn Ile Arg Ala Leu Lys Asp Gly

50 55 60

Tyr Glu Lys Val Glu Gln Asp Leu Pro Ala Phe Glu Asp Leu Glu Gly 65 70 75 80

Ala Ala Arg Ala Leu

<210> 280

<211> 7

<212> PRT

<213> Homo sapiens

<400> 280

Ala Leu Met Arg Leu Gln Asp

<210> 281

<211> 7

<212> PRT

<213> Homo sapiens

<400> 281

Val Glu Ala Gly Gly Ala Thr

<210> 282

<211> 489

<212> PRT

<213> Homo sapiens

<400> 282

Gly Thr Arg Glu Ala Arg Leu Arg Asp Leu Thr Arg Phe Tyr Asp Lys
1 5 10 15

Val Leu Ser Leu His Glu Asp Ser Thr Thr Pro Val Ala Asn Pro Leu 20 25 30

Leu Ala Phe Thr Leu Ile Lys Arg Leu Gln Ser Asp Trp Arg Asn Val
35 40

Val His Ser Leu Glu Ala Ser Glu Asn Ile Arg Ala Leu Lys Asp Gly 50 60

Tyr Glu Lys Val Glu Gln Asp Leu Pro Ala Phe Glu Asp Leu Glu Gly 65 70 75 80

Ala Ala Arg Ala Leu Met Arg Leu Gln Asp Val Tyr Met Leu Asn Val

Lys Gly Leu Ala Arg Gly Val Phe Gln Arg Val Thr Gly Ser Ala Ile $100 \hspace{1cm} 105 \hspace{1cm} 110$

Thr Asp Leu Tyr Ser Pro Lys Arg Leu Phe Ser Leu Thr Gly Asp Asp 115 120 125

- Cys Phe Gln Val Gly Lys Val Ala Tyr Asp Met Gly Asp Tyr Tyr His
- Ala Ile Pro Trp Leu Glu Glu Ala Val Ser Leu Phe Arg Gly Ser Tyr 145 150 155 160
- Gly Glu Trp Lys Thr Glu Asp Glu Ala Ser Leu Glu Asp Ala Leu Asp 165 170 175
- His Leu Ala Phe Ala Tyr Phe Arg Ala Gly Asn Val Ser Cys Ala Leu 180 185 190
- Ser Leu Ser Arg Glu Phe Leu Leu Tyr Ser Pro Asp Asn Lys Arg Met 195 200 205
- Ala Arg Asn Val Leu Lys Tyr Glu Arg Leu Leu Ala Glu Ser Pro Asn 210 215 220
- His Val Val Ala Glu Ala Val Ile Gln Arg Pro Asn Ile Pro His Leu 235 235 240
- Gln Thr Arg Asp Thr Tyr Glu Gly Leu Cys Gln Thr Leu Gly Ser Gln 245 250 255
- Pro Thr Leu Tyr Gln Ile Pro Ser Leu Tyr Cys Ser Tyr Glu Thr Asn 260 265 270
- Ser Asn Ala Tyr Leu Leu Leu Gln Pro Ile Arg Lys Glu Val Ile His 275 280 285
- Leu Glu Pro Tyr Ile Ala Leu Tyr His Asp Phe Val Ser Asp Ser Glu 290 295 300
- Ala Gln Lys Ile Arg Glu Leu Ala Glu Pro Trp Leu Gln Arg Ser Val 305 310 315 320
- Val Ala Ser Gly Glu Lys Gln Leu Gln Val Glu Tyr Arg Ile Ser Lys 325 330 335
- Ser Ala Trp Leu Lys Asp Thr Val Asp Leu Lys Leu Val Thr Leu Asn 340 345 350
- His Arg Ile Ala Ala Leu Thr Gly Leu Asp Val Arg Pro Pro Tyr Ala 355 360 365
- Glu Tyr Leu Gln Val Val Asn Tyr Gly Ile Gly Gly His Tyr Glu Pro 370 375 380
- His Phe Asp His Ala Thr Ser Pro Ser Ser Pro Leu Tyr Arg Met Lys 385 390 400
- Ser Gly Asn Arg Val Ala Thr Phe Met Ile Tyr Leu Ser Ser Val Glu 405 410 415
- Ala Gly Gly Ala Thr Ala Phe Ile Tyr Ala Asn Leu Ser Val Pro Val 420 425 430
- Val Arg Asn Ala Ala Leu Phe Trp Trp Asn Leu His Arg Ser Gly Glu 435 440 445

Gly Asp Ser Asp Thr Leu His Ala Gly Cys Pro Val Leu Val Gly Asp 450 460

Lys Trp Val Ala Asn Lys Trp Ile His Glu Tyr Gly Gln Glu Phe Arg 465 470 470 475

Arg Pro Cys Ser Ser Ser Pro Glu Asp 485

<210> 283

<211> 136

<212> PRT

<213> Homo sapiens

<400> 283

Ile Gln Pro Ser His Ala Ala Leu Leu His Cys Arg Ser Thr Phe Arg
1 10 15

Lys Thr Glu Cys Leu Asp Pro Trp Trp Val Arg Arg Gln Leu Leu Gly 20 25 30

Met Ala Gly Ile Gly Gly Leu Gln Lys Met Lys Ala Pro His Thr Gly 35 40 45

Val Leu His Leu Gly Ser Val Trp Val Phe Leu Gly Pro Phe Leu Leu 50 55 60

Gly Val Gly Tyr Thr Leu Thr Phe Asn Pro Leu Ser Gly Cys Met Ser 65 70 75 80

Thr Val Arg Trp Leu Asn Ser Asn Ile Thr Ala Asn Arg Thr Leu Ser 85 90 95

Arg Ser Val Cys His Val Thr Pro Leu His Arg Ser Leu Ser Pro His 100 105 110

Asp Gly Glu Tyr Leu Arg Gln Met Leu Leu Asn Ser Ser Ser Arg Ala 115 120 125

Gly Glu Ala Gly Ser Trp Gly Tyr

<210> 284

<211> 86

<212> PRT

<213> Homo sapiens

<400> 284

Cys Ser Ser Pro Pro Gly Arg Leu Pro Trp Cys Trp Thr Ala Pro Arg
1 5 10 . 15

Thr Leu Gly Lys His Gly Ser Leu Ile Ser Thr Leu Arg Leu Thr Ala 20 25 30

Pro Leu His Leu Ala Trp Lys Met Met Leu Ser Arg Lys Ala Leu Phe . 35 40 45 Val Leu Leu Asn Thr Pro Val Leu Phe His Ala Leu Glu Gly Arg Leu 50 55 60

Phe Ser Lys Leu Cys His His His Thr Ile Gln Arg Thr Leu Thr Val
65 70 75

Pro Lys Phe Arg Ser Ser

<210> 285

<211> 75

<212> PRT

<213> Homo sapiens

<400> 285

Arg Ser Pro Thr Ser Arg Val Gln Leu Leu Lys Arg Gln Ser Cys Pro

1 5 10 15

Cys Gln Arg Asn Asp Leu Asn Glu Glu Pro Gln His Phe Thr His Tyr
20 25 30

Ala Ile Tyr Asp Phe Ile Val Lys Gly Ser Cys Phe Cys Asn Gly His 35 40 45

Ala Asp Gln Cys Ile Pro Val His Gly Phe Arg Pro Val Lys Ala Pro 50 55 60

Gly Thr Phe His Met Val His Gly Lys Cys Met

<210> 286

<211> 296

<212> PRT

<213> Homo sapiens

<400> 286

His Asn Thr Ala Gly Ser His Cys Gln His Cys Ala Pro Leu Tyr Asn

1 5 10 15

Asp Arg Pro Trp Glu Ala Ala Asp Gly Lys Thr Gly Ala Pro Asn Glu 20 25 30

Cys Arg Thr Cys Lys Cys Asn Gly His Ala Asp Thr Cys His Phe Asp

Val Asn Val Trp Glu Ala Ser Gly Asn Arg Ser Gly Gly Val Cys Asp
50 55 60

Asp Cys Gln His Asn Thr Glu Gly Gln Tyr Cys Gln Arg Cys Lys Pro
65 70 75 80

Gly Phe Tyr Arg Asp Leu Arg Arg Pro Phe Ser Ala Pro Asp Ala Cys
85 90

Lys Pro Cys Ser Cys His Pro Val Gly Ser Ala Val Leu Pro Ala Asn 100 105 110 Ser Val Thr Phe Cys Asp Pro Ser Asn Gly Asp Cys Pro Cys Lys Pro 115 120 125

Gly Val Ala Gly Arg Arg Cys Asp Arg Cys Met Val Gly Tyr Trp Gly
130 135 140

Phe Gly Asp Tyr Gly Cys Arg Pro Cys Asp Cys Ala Gly Ser Cys Asp 145 150 155 160

Pro Ile Thr Gly Asp Cys Ile Ser Ser His Thr Asp Ile Asp Trp Tyr 165 170 175

His Glu Val Pro Asp Phe Arg Pro Val His Asn Lys Ser Glu Pro Ala 180 185 190

Trp Glu Trp Glu Asp Ala Gln Gly Phe Ser Ala Leu Leu His Ser Gly 195 200 205

Lys Cys Glu Cys Lys Glu Gln Thr Leu Gly Asn Ala Lys Ala Phe Cys 210 215 220

Gly Met Lys Tyr Ser Tyr Val Leu Lys Ile Lys Ile Leu Ser Ala His 225 230 235 . 240

Asp Lys Gly Thr His Val Glu Val Asn Val Lys Ile Lys Lys Val Leu 245 250 255

Lys Ser Thr Lys Leu Lys Ile Phe Arg Gly Lys Ala Asn Ile Ile Ser 260 . 265 270

Arg Ile Met Asp Gly Gln Arg Met His Leu Ser Asn Pro Gln Ser Trp 275 280 285

Phe Gly Ile Pro Cys Ser Arg Thr 290 295

<210> 287

<211> 37

<212> PRT

<213> Homo sapiens

<400> 287

Cys Asp Asp Cys Gln His Asn Thr Glu Gly Gln Tyr Cys Gln Arg Cys

1 10 15

Lys Pro Gly Phe Tyr Arg Asp Leu Arg Arg Pro Phe Ser Ala Pro Asp 20 25 30

Ala Cys Lys Pro Cys 35

<210> 288

<211> 36

<212> PRT

<213> Homo sapiens

```
<400> 288
```

Cys Pro Cys Lys Pro Gly Val Ala Gly Arg Arg Cys Asp Arg Cys Met

1 5 10 15

Val Gly Tyr Trp Gly Phe Gly Asp Tyr Gly Cys Arg Pro Cys Asp Cys
20 25 30

Ala Gly Ser Cys 35

<210> 289

<211> 66

<212> PRT

<213> Homo sapiens

<400> 289

Asn Ile Ser Ser Gln Tyr Cys Ile Leu Lys Ser Leu Glu Met Met Ile

1 10 15

Ser Gly Leu Lys Leu Leu Val Leu Phe Leu Lys Phe Ala Pro Glu Asn 20 25 25 30

Tyr Cys Leu Ser Thr Glu Thr Leu Gln Met Pro Asn Arg His Leu Arg 35 40 45

Leu Ser Lys Ala Thr Cys Tyr Leu Met Lys Cys Leu Leu Pro Ser Tyr 50 55 60

Phe Glu 65

<210> 290

<211> 88

<212> PRT

<213> Homo sapiens

<400> 290

Pro Ile Glu Gly Thr Pro Ala Gly Thr Gly Pro Glu Phe Pro Gly Arg

1 5 10 15

Pro Thr Arg Pro Gln Arg Met Arg Ser Leu Ile Ser Ser His Pro Cys
20 25 30

Gln His Leu Leu Leu Leu Leu Leu Leu Leu Phe Leu Ile Leu Ala Ile

Leu Val Asp Val Lys Trp Tyr Leu Val Leu Phe Ile Cys Ile Ser Leu 50 55 60

Met Thr Ser Asp Val Glu His Leu Phe Met Cys Leu Leu Ala Ile Arg 65 70 75 80

Ile Ser Ser Trp Arg Asn Val Tyr

<210> 291

195

PCT/US99/17130

```
<211> 60
 <212> PRT
 <213> Homo sapiens
 <400> 291
 Asn Trp Val Pro Thr Cys Leu Cys Pro Ser Ala Pro Cys Ser Phe His
                                       10
 Leu Leu Ser Arg Phe Lys Cys Leu Phe Ser Pro Gln Arg Leu Thr Asp
 Ile Phe Arg Arg Tyr Asp Thr Asp Gln Asp Gly Trp Ile Gln Val Ser
          35
                               40
 Tyr Glu Gln Tyr Leu Ser Met Val Phe Ser Ile Val
                          55
 <210> 292
 <211> 33
 <212> PRT
 <213> Homo sapiens
<400> 292
Gln Arg Leu Thr Asp Ile Phe Arg Arg Tyr Asp Thr Asp Gln Asp Gly
                                      10
Trp Ile Gln Val Ser Tyr Glu Gln Tyr Leu Ser Met Val Phe Ser Ile
                                  25
Val
<210> 293
<211> 73
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (38)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (48)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (54)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
```

<223> Xaa equals any of the naturally occurring L-amino acids

<222> (55)

```
<220>
```

<221> SITE

<222> (68)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 293

Met Phe Tyr Lys Leu Thr Leu Ile Leu Cys Glu Leu Ser Val Ala Gly

1 5 10 15

Val Thr Gln Ala Ala Ser Gln Arg Pro Leu Gln Arg Leu Pro Arg His 20 25 30

Ile Cys Ser Gln Arg Xaa Pro Pro Gly Arg Cys Leu Leu Lys Ala Xaa 35 40 . 45

Leu Gln Thr Trp Xaa Xaa Pro Asp Lys Pro Ile Pro Arg Leu Ser 50 55 60

Pro Pro Leu Xaa Ser Asp Pro Lys Arg 65 70

<210> 294

<211> 95

<212> PRT

<213> Homo sapiens

<400> 294

Thr Ser Ser Pro Val Phe Ser Phe Cys Ser Met Ala Val Arg Glu Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Asp His Leu Gln Arg Val Ser Leu Pro Arg Tyr Asn Val Ser Ala Ser 20 25 30

Leu Gln Trp Leu Pro Cys His Arg Ile Val Leu Gln Pro Trp His Met 35 40 45

Cys Ala Met Trp Glu Leu Gly Gln Val Leu Phe His Pro Val Ala Pro 50 55 60

Arg Glu Gly Ala Ala Pro Ser Pro Val Ser Thr Leu Thr Trp Pro Ser 65 70 75 80

Ser Cys Ser His Ser Glu Ser Thr Met Glu Leu Glu Leu Gln Phe 85 90 95

<210> 295

<211> 16

<212> PRT

<213> Homo sapiens

<400> 295

Met Ala Val Arg Glu Pro Asp His Leu Gln Arg Val Ser Leu Pro Arg

1 10 15

```
<210> 296
<211> 7
<212> PRT
<213> Homo sapiens
<400> 296
Leu Pro Cys His Arg Ile Val
<210> 297
<211> 15
<212> PRT
<213> Homo sapiens
<400> 297
Ser Leu Gln Trp Leu Pro Cys His Arg Ile Val Leu Gln Pro Trp
<210> 298
<211> 454
<212> PRT
<213> Homo sapiens
<400> 298
Cys Phe Lys Arg Lys Pro Lys Arg Glu His Cys Ser Cys Pro Ile Thr
                                     10
Tyr Gln Ser Leu Gly Asp Ile Leu Asn Ala Ser Phe Phe Ser Lys Arg
Lys Gly Met Gln Glu Val Lys Leu Asn Ser Tyr Val Val Ser Gly Thr
                            40
Ile Gly Leu Lys Glu Lys Ile Ser Leu Ser Glu Pro Val Phe Leu Thr
Phe Arg His Asn Gln Pro Gly Asp Lys Arg Thr Lys His Ile Cys Val
Tyr Trp Glu Gly Ser Glu Gly Gly Arg Trp Ser Thr Glu Gly Cys Ser
His Val His Ser Asn Gly Ser Tyr Thr Lys Cys Lys Cys Phe His Leu
                               105
Ser Ser Phe Ala Val Leu Val Ala Leu Ala Pro Lys Glu Asp Pro Val
                            120
Leu Thr Val Ile Thr Gln Val Gly Leu Thr Ile Ser Leu Leu Cys Leu
Phe Leu Ala Ile Leu Thr Phe Leu Leu Cys Arg Pro Ile Gln Asn Thr
                    150
Ser Thr Ser Leu His Leu Glu Leu Ser Leu Cys Leu Phe Leu Ala His
                165
                                   170
```

- Leu Leu Phe Leu Thr Gly Ile Asn Arg Thr Glu Pro Glu Val Leu Cys 180 185 190
- Ser Ile Ile Ala Gly Leu Leu His Phe Leu Tyr Leu Ala Cys Phe Thr 195 200 205
- Trp Met Leu Leu Glu Gly Leu His Leu Phe Leu Thr Val Arg Asn Leu 210 215 220
- Lys Val Ala Asn Tyr Thr Ser Thr Gly Arg Phe Lys Lys Arg Phe Met 225
- Tyr Pro Val Gly Tyr Gly Ile Pro Ala Val Ile Ile Ala Val Ser Ala 245 250 255
- Ile Val Gly Pro Gln Asn Tyr Gly Thr Phe Thr His Cys Trp Leu Lys 260 265 270
- Leu Asp Lys Gly Phe Ile Trp Ser Phe Met Gly Pro Val Ala Val Ile 275 280 285
- Ile Leu Ile Asn Leu Val Phe Tyr Phe Gln Val Leu Trp Ile Leu Arg 290 295 300
- Ser Lys Leu Ser Ser Leu Asn Lys Glu Val Ser Thr Ile Gln Asp Thr 305 310 315 320
- Ser Trp Gly Leu Gly Phe Phe Met Val Glu Glu Val Gly Lys Thr Ile 340 345 350
- Gly Ser Ile Ile Ala Tyr Ser Phe Thr Ile Ile Asn Thr Leu Gln Gly 355 360 365
- Val Leu Leu Phe Val Val His Cys Leu Leu Asn Arg Gln Val Arg Met 370 380
- Glu Tyr Lys Lys Trp Phe Ser Gly Met Arg Lys Gly Val Glu Thr Glu 385 390 395 400
- Ser Thr Glu Met Ser Arg Ser Thr Thr Gln Thr Lys Thr Glu Glu Val 405 410 415
- Gly Lys Ser Ser Glu Ile Phe His Lys Gly Gly Thr Ala Ser Ser Ser 420 425 430
- Ala Glu Ser Thr Lys Gln Pro Gln Pro Gln Val His Leu Val Ser Ala 435
- Ala Trp Leu Lys Met Asn 450

<210> 299

<211> 101

<212> PRT

<213> Homo sapiens

<400> 299

Phe Phe Trp Lys Glu Asn Leu Arg Arg Asn Gly Ser Arg Glu Asp Phe 1 5 10 15

Ala Arg Arg Ala Thr Gln Leu Ile Gln Ser Val Glu Leu Ser Ile Trp
20 25 30

Asn Ala Ser Phe Ala Ser Pro Gly Lys Gly Gln Ile Ser Glu Phe Asp 35 40 45

Ile Val Tyr Glu Thr Lys Arg Cys Asn Glu Thr Arg Glu Asn Ala Phe 50 55 60

Leu Glu Ala Gly Asn Asn Thr Met Asp Ile Asn Cys Ala Asp Ala Leu
65 70 75 80

Lys Gly Asn Leu Arg Glu Ser Thr Ala Val Ala Leu Ser Leu Ile Asn 85 90 95

Leu Leu Gly Ile Phe 100

<210> 300

<211> 27

<212> PRT

<213> Homo sapiens

<400> 300

Asp Ile Asn Glu Cys Glu Thr Gly Leu Ala Lys Cys Lys Tyr Lys Ala 1 5 10 15

Tyr Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys 20 25

<210> 301

<211> 12

<212> PRT

<213> Homo sapiens

<400> 301

Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys Ser Cys 1 5 10

<210> 302

<211> 331

<212> PRT

<213> Homo sapiens

<400> 302

Ala Leu Cys Pro His Pro His Leu Ile Leu Asn Val Thr Val Ser Pro 1 5 10 15

Ala Pro Ser Cys Arg His Val Lys Lys Val Val Ala Ser Pro Ser Pro 20 25 30

- Ser Thr Thr Met Ile Ala Met Asp Ala Pro His Ser Lys Ala Ala Leu 35 40 45
- Asp Ser Ile Asn Glu Leu Pro Glu Asn Ile Leu Leu Glu Leu Phe Thr 50 55 60
- His Val Pro Ala Arg Gln Leu Leu Leu Asn Cys Arg Leu Val Cys Ser 65 70 75 80
- Leu Trp Arg Asp Leu Ile Asp Leu Met Thr Leu Trp Lys Arg Lys Cys
 85 90 95
- Leu Arg Glu Gly Phe Ile Thr Lys Asp Trp Asp Gln Pro Val Ala Asp
 100 105 110
- Trp Lys Ile Phe Tyr Phe Leu Arg Ser Leu His Arg Asn Leu Leu Arg 115 120 125
- Asn Pro Cys Ala Glu Glu Asp Met Phe Ala Trp Gln Ile Asp Phe Asn 130 135 140
- Gly Gly Asp Arg Trp Lys Val Glu Ser Leu Pro Gly Ala His Gly Thr 150 150 155
- Asp Phe Pro Asp Pro Lys Val Lys Lys Tyr Phe Val Thr Ser Tyr Glu 165 170 175
- Met Cys Leu Lys Ser Gln Leu Val Asp Leu Val Ala Glu Gly Tyr Trp 180 185 190
- Glu Glu Leu Leu Asp Thr Phe Arg Pro Asp Ile Val Val Lys Asp Trp
 195 200 205
- Phe Ala Ala Arg Ala Asp Cys Gly Cys Thr Tyr Gln Leu Lys Val Gln 210 215 220
- Leu Ala Ser Ala Asp Tyr Phe Val Leu Ala Ser Phe Glu Pro Pro 225 230 235. 240
- Val Thr Ile Gln Gln Trp Asn Asn Ala Thr Trp Thr Glu Val Ser Tyr 245 250 255
- Thr Phe Ser Asp Tyr Pro Arg Gly Val Arg Tyr Ile Leu Phe Gln His 260 265 270
- Gly Gly Arg Asp Thr Gln Tyr Trp Ala Gly Trp Tyr Gly Pro Arg Val 275 280 285
- Thr Asn Ser Ser Ile Val Val Ser Pro Lys Met Thr Arg Asn Gln Ala 290 295 300
- Ser Ser Glu Ala Gln Pro Gly Gln Lys His Gly Gln Glu Glu Ala Ala 305 310 315 320
- Gln Ser Pro Tyr Arg Ala Val Val Gln Ile Phe 325 330

<210> 303

<211> 328

<212> PRT

<213> Homo sapiens

<400> 303

Arg Gln Arg Ser Trp Asn Pro Gly Thr Asn Cys Tyr His Pro Asn Met $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Pro Asp Ala Phe Leu Thr Cys Glu Thr Val Ile Phe Ala Trp Ala Ile 20 25 30

Gly Gly Glu Gly Phe Ser Tyr Pro Pro His Val Gly Leu Ser Leu Gly 35 40 45

Thr Pro Leu Asp Pro His Tyr Val Leu Leu Glu Val His Tyr Asp Asn 50 55 60

Pro Thr Tyr Glu Glu Gly Leu Ile Asp Asn Ser Gly Leu Arg Leu Phe 65 70 75 80

Tyr Thr Met Asp Ile Arg Lys Tyr Asp Ala Gly Val Ile Glu Ala Gly
85 90 95

Leu Trp Val Ser Leu Phe His Thr Ile Pro Pro Gly Met Pro Glu Phe 100 105 110

Gln Ser Glu Gly His Cys Thr Leu Glu Cys Leu Glu Glu Ala Leu Glu 115 120 125

Ala Glu Lys Pro Ser Gly Ile His Val Phe Ala Val Leu His Ala 130 135 140

His Leu Ala Gly Arg Gly Ile Arg Leu Arg His Phe Arg Lys Gly Lys 145 150 155 160

Glu Met Lys Leu Leu Ala Tyr Asp Asp Phe Asp Phe Asn Phe Gln
165 170 175

Glu Phe Gln Tyr Leu Lys Glu Glu Gln Thr Ile Leu Pro Gly Asp Asn 180 185 190

Leu Ile Thr Glu Cys Arg Tyr Asn Thr Lys Asp Arg Ala Glu Met Thr 195 200 205

Trp Gly Gly Leu Ser Thr Arg Ser Glu Met Cys Leu Ser Tyr Leu Leu 210 215 220

Tyr Tyr Pro Arg Ile Asn Leu Thr Arg Cys Ala Ser Ile Pro Asp Ile 225 230 235 240

Met Glu Gln Leu Gln Phe Ile Gly Val Lys Glu Ile Tyr Arg Pro Val 245 250 255

Thr Thr Trp Pro Phe Ile Ile Lys Ser Pro Lys Gln Tyr Lys Asn Leu 260 265 270

Ser Phe Met Asp Ala Met Asn Lys Phe Lys Trp Thr Lys Lys Glu Gly 275 280 285

Leu Ser Phe Asn Lys Leu Val Leu Ser Leu Pro Val Asn Val Arg Cys 290 295 300

Ser Lys Thr Asp Asn Ala Glu Trp Ser Ile Pro Arg Asn Asp Ser Ile 305 310 315 320

Thr Ser Arg Tyr Arg Lys Thr Leu 325

<210> 304

<211> 272

<212> PRT

<213> Homo sapiens

<400> 304

Met Cys Cys Trp Pro Leu Leu Leu Trp Gly Leu Leu Pro Gly Thr
1 5 10 15

Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu Leu Asp 20 25 30

Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly Ser Gln Ile 35 40 45

Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val Gly Phe Gly Phe 50 55 60

Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile Val Val Gly Gly Val 65 70 75 80

Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr Phe Thr Asn Ala Asn Arg 85 90 95

Glu Leu Lys Lys Asp Ala Gln Gln Asp Tyr His Leu Glu Tyr Ala Met 100 105 110

Glu Asn Ser Thr His Thr Ile Ile Glu Phe Thr Arg Glu Leu His Thr 115 120 125

Cys Asp Ile Asn Asp Lys Ser Ile Thr Asp Ser Thr Val Arg Val Ile 130 135 140

Trp Ala Tyr His His Glu Asp Ala Gly Glu Ala Gly Pro Lys Tyr His 145 150 155 160

Asp Ser Asn Arg Gly Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys
165 170 175

Thr Ser Val Leu Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln
180 185 190

Asp Val Pro Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe
195 200 205

Lys Ile Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro 210 215 220

Val Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr 225 230 235 240

Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Pro Gly Ile Arg Ala Arg 245 250 255

Ile Ala Ile Thr Pro Thr Cys Pro Met His Ser Ser Pro Val Lys Leu 260 265 270

<210> 305

<211> 207

<212> PRT

<213> Homo sapiens

<400> 305

Thr Gly Thr Phe Trp Ser Pro Arg Ser Gln Arg Arg Gly Cys Cys Gly

1 10 15

Arg Arg Ala Pro Arg Pro Glu Ala Met Glu Asn Gly Ala Val Tyr Ser 20 25 30

Pro Thr Thr Glu Glu Asp Pro Gly Pro Ala Arg Gly Pro Arg Ser Gly 35 40

Leu Ala Ala Tyr Phe Phe Met Gly Arg Leu Pro Leu Leu Arg Arg Val
50 60

Leu Lys Gly Leu Gln Leu Leu Leu Ser Leu Leu Ala Phe Ile Cys Glu 65 70 75 80

Glu Val Val Ser Gln Cys Thr Leu Cys Gly Gly Leu Tyr Phe Phe Glu 85 90 95

Phe Val Ser Cys Ser Ala Phe Leu Leu Ser Leu Leu Ile Leu Ile Val 100 105 110

Tyr Cys Thr Pro Phe Tyr Glu Arg Val Asp Thr Thr Lys Val Lys Ser 115 120 125

Ser Asp Phe Tyr Ile Thr Leu Gly Thr Gly Cys Val Phe Leu Leu Ala 130 135 140

Ser Ile Ile Phe Val Ser Thr His Asp Arg Thr Ser Ala Glu Ile Ala 145 150 155 160

Ala Ile Val Phe Gly Phe Ile Ala Ser Phe Met Phe Leu Leu Asp Phe 165 170 175

Ile Thr Met Leu Tyr Glu Lys Arg Gln Glu Ser Gln Leu Arg Lys Pro 180 185 190

Glu Asn Thr Thr Arg Ala Glu Ala Leu Thr Glu Pro Leu Asn Ala 195 200 205 <210> 306

<211> 135 <212> PRT

<213> Homo sapiens

<400> 306

Ala Ser Ala Pro Arg Val Met Arg Gly His Leu Ala Gly Phe Pro Ala 1 5 10 15

Leu Ser Gly Leu Ala Ser Val Cys Leu Trp Ala Thr Phe Ser Ala Gln
20 25 30

Leu Pro Gly Pro Val Ala Ala Thr Ser Trp Thr Pro Ala Pro Leu Gly 35 40 45

Cys Ser Ala Ala Arg Ser Gly Pro Glu Lys Arg Leu Gly Thr Ala Ala 50 55 60

Pro Gly Ser Ala Ala Ser Leu Ala Gln Ala Gly Pro Gly Ala Pro Cys 65 70 75 80

Arg Val Leu Pro Val Asp Pro Ala Pro Ala Ala Leu Asn Val Arg Glu
85 90 95

Pro Gly Trp Leu Gly Gly Leu Phe Asp Gly Ala Leu Leu Gln Val Leu
100 105 110

Leu Asn Phe Leu Arg Lys Ser Thr Asp Val Leu Met Asp Thr Arg Glu 115 120 125

Ala Glu Ser Leu Glu Val Glu 130 135

<210> 307

<211> 188

<212> PRT

<213> Homo sapiens

<400> 307

Asn Lys Leu His Ser Phe Pro Val Phe Leu Ser Gln Leu Leu Leu Asp

1 10 15

Arg Gln Leu Leu His Ala Pro Gln Thr Leu Pro Thr Pro His Cys Gly
20 25 30

Gly Ser Ser Arg Pro Gly Pro Ser His Pro Pro Trp Leu Leu Ile Gln
35 40 45

Leu Pro Cys Val His Val Ala Leu Trp Gln Met Leu Arg Asp Phe Ser 50 55 60

Asp Ser Arg Ile Thr Pro Ser Thr Leu Thr Thr Gln Pro Ala Ala Gln 65 70 75 80

Thr Ala Ala Pro Ala Lys Asp Gln Glu Ser Asp Ile Val Gly Glu
85 90 95

Gly Ile Leu Cys Asp Ile Ala Phe Leu Gln Glu Asp His Pro Leu Gly

100 105 110

Val Gly Gly Ala Ser Ala Pro Ser Ser Arg Arg Glu Leu Ser Arg Arg 115 120 125

Gly Val His Thr Gln Thr Leu Pro Glu Asp Gly Thr Leu His Gly Thr 130 135 140

Pro Ser Ser Ser Phe Asp Cys Gly Ile Lys Tyr Ile Ile Ser Trp Pro 145 150 155 160

Leu Ala Pro Gly Cys Asp Leu Pro Ser Leu Glu Leu Ser Leu Val Cys 165 170 175

Lys Gly Val Ser Ser Cys Met Gly Phe Ala Ala Gly
180 185

<210> 308

<211> 78

<212> PRT

<213> Homo sapiens

<400> 308

Pro Gly Arg Pro Thr Arg Pro Thr Lys Asn Lys Val Cys Val Cys Leu

1 10 15

Gly Met Leu Phe Trp Ala Tyr Pro Ile Cys Val Phe Ile Asp Ser Leu 20 25 30

Ser Cys Gln Pro Cys Leu Trp Ser Thr Gly Ala Thr Ser His Phe Asn 35 40

Ser Pro Thr Thr Ser Pro Leu Phe Thr Leu Phe Met Pro Cys Ala Leu 50 60

Ala Pro Asn Pro Phe Thr Gln Leu Gly Lys Leu Asp Asp Arg 65 70 75

<210> 309

<211> 10

<212> PRT

<213> Homo sapiens

<400> 309

Pro Val Asp Leu Thr Lys Thr Arg Leu Gln
1 5 10

<210> 310

<211> 10

<212> PRT

<213> Homo sapiens

<400> 310

Pro Thr Asp Val Leu Lys Ile Arg Met Gln 1 5 10

<2 <2	210> 211> 212> 213>	313 PRT	o sag	oien:	5	•									-
<2 <2	20> 21> 22> 23>	(117	7)	ıls a	iny o	f th	ne na	tura	ılly	occu	rrin	ıg L-	amin	o ac	ids:
	00>														
Me			e Gl	y Se	r Th 5	r Il	e Se	r Pr	o Th	ır Se 0	r Th	r Hi	s Al		r Pro
Se	r Le	u Gl	y Ph 2	е Су 0	ѕ Су	s Se	r Tr	p Le 2		u Gl	u As	p Le	u Gl		u Glr
Let	а Ту	r Cy 3	s Se 5	r Al	a Ph	e Gl	u Gla	u Al	a Al	a Le	u Th	r Ar		g Il	e Cys
Asr	Pro	o Th	r Se	r Cy	s Trp	Le 5	u Pro) Le	u As	p Me	Gli 60		ı Leı	ı Hi	s Arg
Glr 65	ı Val	l Le	u Ala	a Le	Glr 70	Th:	r Glr	n Arg	y Vai	l Let 79	ı Let	ı Gly	/ Met	Tr	Leu 80
Arg	Arg	g Ala	a Tri	Ası 89	Thr	Tr	Val	Ser	Pro 90		Arg	y Val	. Ala	Pro 95	Gly
Ser	Arg	g Cys	100	ı Leı)	Thr	Ala	Ser	His 105	Pro	Суз	Thr	Glu	Lys 110		Arg
Lys	Ala	Ser 115	Ala	Xaa	Gln	Arg	Asn 120	Leu	Gly	Tyr	Pro	Leu 125		Met	Leu
Cys	Leu 130	Leu	Val	Leu	Thr	Gly 135	Leu	Ser	Val	Leu	Ile 140	Val	Ala	Ile	His
Ile 145	Leu	Glu	Leu	Leu	Ile 150	Asp	Glu	Ala	Ala	Met 155	Pro	Arg	Gly	Met	Gln 160
Gly	Thr	Ser	Ļeu	Gly 165	Gln	Val	Ser	Phe	Ser 170	Lys	Leu	Gly	Ser	Phe 175	Gly
Ala	Val	Ile	Gln 180	Val	Val	Leu	Ile	Phe 185	Tyr	Leu	Met	Val	Ser 190	Ser	Val
Val	Gly	Phe 195	Tyr	Ser	Ser	Pro	Leu 200	Phe	Arg	Ser	Leu	Arg 205	Pro	Arg	Trp
His	Asp 210	Thr	Ala	Met	Thr	Gln 215	Ile	Ile	Gly	Asn	Cys 220	Val	Cys	Leu	Leu
Val 225	Leu	Ser	Ser	Ala	Leu 230	Pro	Val	Phe	Ser	Arg 235	Thr	Leu	Gly	Leu	Thr 240

Arg Phe Asp Leu Leu Gly Asp Phe Gly Arg Phe Asn Trp Leu Gly Asn 245 250 255

- ----

Phe Tyr Ile Val Phe Leu Tyr Asn Ala Ala Phe Ala Gly Leu Thr Thr 260 265 270

Leu Cys Leu Val Lys Thr Phe Thr Ala Ala Val Arg Ala Glu Leu Ile 275 280 285

Arg Ala Phe Gly Leu Asp Arg Leu Pro Leu Pro Val Ser Gly Phe Pro 290 295 300

Gln Ala Ser Arg Lys Thr Gln His Gln 305 310

<210> 312

<211> 92

<212> PRT

<213> Homo sapiens

<400> 312

Leu Cys Val Cys Leu Val Tyr Leu Cys Met Tyr Gly Val Cys Leu Cys 1 5 10 15

Val Ile Val Cys Val Ser Gly Val Ser Leu Cys Leu Tyr Val Trp Gly
20 25 30

Val Ser Val Cys Asp Cys Val Ser Val Phe Met Cys Val Cys Leu Cys 35 40

Val Ile Phe Cys Val Tyr Gly Lys Pro Arg Thr Glu His Tyr His Ser 50 55 60

Pro His Leu Ala Lys Gln Lys Ala Phe Arg Glu Met Cys Gly Arg His 65 70 75 80

Asp Val Ser Ala Ala Gly Ile Phe Gln Ser Tyr Val

<210> 313

<211> 207

<212> PRT

<213> Homo sapiens

<400> 313

Gly His Met Pro Tyr Gly Trp Leu Thr Glu Ile Arg Ala Val Tyr Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Ala Phe Asp Lys Asn Asn Pro Ser Asn Lys Leu Val Ser Thr Ser Asn 20

Thr Val Thr Ala Ala His Ile Lys Lys Phe Thr Phe Val Cys Met Ala 35 40 45

Leu Ser Leu Thr Leu Cys Phe Val Met Phe Trp Thr Pro Asn Val Ser 50 55 60

Glu Lys Ile Leu Ile Asp Ile Ile Gly Val Asp Phe Ala Phe Ala Glu 65 70 75 80 Leu Cys Val Val Pro Leu Arg Ile Phe Ser Phe Phe Pro Val Pro Val 85 90 95

Thr Val Arg Ala His Leu Thr Gly Trp Leu Met Thr Leu Lys Lys Thr

Phe Val Leu Ala Pro Ser Ser Val Leu Arg Ile Ile Val Leu Ile Ala 115 120 125

Ser Leu Val Val Leu Pro Tyr Leu Gly Val His Gly Ala Thr Leu Gly 130 140

Val Gly Ser Leu Leu Ala Gly Phe Val Gly Glu Ser Thr Met Val Ala 145 150 155 160

Ile Ala Ala Cys Tyr Val Tyr Arg Lys Gln Lys Lys Lys Met Glu Asn 165 170 175

Glu Ser Ala Thr Glu Gly Glu Asp Ser Ala Met Thr Asp Met Pro Pro 180 185 190

Thr Glu Glu Val Thr Asp Ile Val Glu Met Arg Glu Glu Asn Glu
195 200 205

<210> 314

<211> 114

<212> PRT

<213> Homo sapiens

<400> 314

Glu Pro Leu Leu Ile Pro Ile Leu Ser Leu Tyr Met Gly Ala Leu Val 20 25 30

Arg Cys Thr Thr Leu Cys Leu Gly Tyr Tyr Lys Asn Ile His Asp Ile 35 40 45

Ile Pro Asp Arg Ser Gly Pro Glu Leu Gly Gly Asp Ala Thr Ile Arg 50 55 60

Lys Met Leu Ser Phe Trp Trp Pro Leu Ala Leu Ile Leu Ala Thr Gln 65 70 75 80

Arg Ile Ser Arg Pro Ile Val Asn Leu Phe Val Ser Arg Asp Leu Gly

Gly Ser Ser Ala Ala Thr Glu Ala Val Ala Ile Leu Thr Ala Thr Tyr 100 105 110

Pro Val

<210> 315

<211> 115

<212> PRT

<213> Homo sapiens

<400> 315

Arg Cys Cys Cys Arg Gly Cys Ser Cys Arg Ala Arg Leu Cys Pro Pro 1 5 10 15

Ala Arg Ser Thr Ala Val Ala Pro Glu Cys Arg Gly Ala His Pro Ser 20 25 30

Arg Ala Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val 35 40 45

Leu Leu Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln 50 55 60

Pro Val Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr 65 70 75 80

Phe His Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala 85 90 95

Cys Arg Arg Gly Trp Arg Pro Ala Ser Gln His Arg Val Leu Lys Met 100 105 110

Asn Arg Asn 115

<210> 316

<211> 81

<212> PRT

<213> Homo sapiens

<400> 316

Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val Leu Leu $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln Pro Val 20 25 30

Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr Phe His $35 \ 40 \ 45$

Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala Cys Arg 50 55 60

Arg Gly Trp Arg Pro Ala Ser Gln His Arg Val Leu Lys Met Asn Arg 65 70 75 80

Asn

<210> 317

<211> 290

<212> PRT

<213> Homo sapiens

:--

<	4	0	0	>	3	1	7
<	4	υ	V	>	3	1	7

- Ile Arg His Glu Gln Gln Gly Glu Glu Asp Asp Glu His Ala Arg Pro

 1 5 10 15
- Leu Ala Glu Ser Leu Leu Leu Ala Ile Ala Asp Leu Leu Phe Cys Pro 20 25 30
- Asp Phe Thr Val Gln Ser His Arg Arg Ser Thr Val Asp Ser Ala Glu
 35 40 45
- Asp Val His Ser Leu Asp Ser Cys Glu Tyr Ile Trp Glu Ala Gly Val
 50 55 60
- Gly Phe Ala His Ser Pro Gln Pro Asn Tyr Ile His Asp Met Asn Arg
 65 70 75
- Met Glu Leu Leu Lys Leu Leu Leu Thr Cys Phe Ser Glu Ala Met Tyr
- Leu Pro Pro Ala Pro Glu Ser Gly Ser Thr Asn Pro Trp Val Gln Phe
- Phe Cys Ser Thr Glu Asn Arg His Ala Leu Pro Leu Phe Thr Ser Leu 115 120 125
- Leu Asn Thr Val Cys Ala Tyr Asp Pro Val Gly Tyr Gly Ile Pro Tyr 130 140
- Asn His Leu Leu Phe Ser Asp Tyr Arg Glu Pro Leu Val Glu Glu Ala 145 150 155 160
- Ala Gln Val Leu Ile Val Thr Leu Asp His Asp Ser Ala Ser Ser Ala 165 170 175
- Ser Pro Thr Val Asp Gly Thr Thr Thr Gly Thr Ala Met Asp Asp Ala 180 185 190
- Asp Pro Pro Gly Pro Glu Asn Leu Phe Val Asn Tyr Leu Ser Arg Ile 195
- His Arg Glu Glu Asp'Phe Gln Phe Ile Leu Lys Gly Ile Ala Arg Leu 210 215 220
- Leu Ser Asn Pro Leu Leu Gln Thr Tyr Leu Pro Asn Ser Thr Lys Lys 235 240
- Asp Pro Val Pro Pro Gly Ala Ala Ser Ser Leu Leu Glu Ala Leu Arg
- Leu Gln Gln Glu Ile Pro Leu Leu Arg Ala Glu Glu Gln Arg Arg Pro 260 265 270
- Arg His Pro Cys Pro His Pro Leu Leu Pro Gln Arg Cys Pro Gly Arg

Ser Val

290

<210> 318 <211> 318 <212> PRT <213> Homo sapiens <400> 318 5 Leu Asp Pro Ser Lys Ser Ser Val Gly Ser Tyr Phe His Thr Met Val Glu Ser Leu Val Gly Trp Gly Tyr Thr Arg Gly Glu Asp Val Arg Gly Ala Pro Tyr Asp Trp Arg Arg Ala Pro Asn Glu Asn Gly Pro Tyr Phe

Arg Leu Val Tyr Asn Lys Thr Ser Arg Ala Thr Gln Phe Pro Asp Gly 10 Val Asp Val Arg Val Pro Gly Phe Gly Lys Thr Phe Ser Leu Glu Phe

90

Leu Ala Leu Arg Glu Met Ile Glu Glu Met Tyr Gln Leu Tyr Gly Gly

Pro Val Val Leu Val Ala His Ser Met Gly Asn Met Tyr Thr Leu Tyr

Phe Leu Gln Arg Gln Pro Gln Ala Trp Lys Asp Lys Tyr Ile Arg Ala

Phe Val Ser Leu Gly Ala Pro Trp Gly Gly Val Ala Lys Thr Leu Arg

Val Leu Ala Ser Gly Asp Asn Asn Arg Ile Pro Val Ile Gly Pro Leu

Lys Ile Arg Glu Gln Gln Arg Ser Ala Val Ser Thr Ser Trp Leu Leu 170

Pro Tyr Asn Tyr Thr Trp Ser Pro Glu Lys Val Phe Val Gln Thr Pro 185

Thr Ile Asn Tyr Thr Leu Arg Asp Tyr Arg Lys Phe Phe Gln Asp Ile 200

Gly Phe Glu Asp Gly Trp Leu Met Arg Gln Asp Thr Glu Gly Leu Val

Glu Ala Thr Met Pro Pro Gly Val Gln Leu His Cys Leu Tyr Gly Thr 235

Gly Val Pro Thr Pro Asp Ser Phe Tyr Tyr Glu Ser Phe Pro Asp Arg 245

Asp Pro Lys Ile Cys Phe Gly Asp Gly Asp Gly Thr Val Asn Leu Lys

Ser Ala Leu Gln Cys Gln Ala Trp Gln Ser Arg Gln Glu His Gln Val

- Leu Cln Glu Leu Pro Gly Ser Glu His Ile Glu Met Leu Ala Asn 290 295 300
- Ala Thr Thr Leu Ala Tyr Leu Lys Arg Val Leu Leu Gly Pro 305 315
- <210> 319
- <211> 362
- <212> PRT
- <213> Homo sapiens
- <400> 319
- Met Asn Lys Glu Asp Lys Val Trp Asn Asp Cys Lys Gly Val Asn Lys

 1 10 15
- Leu Thr Asn Leu Glu Glu Gln Tyr Ile Ile Leu Ile Phe Gln Asn Gly
 20 25 30
- Leu Asp Pro Pro Ala Asn Met Val Phe Glu Ser Ile Ile Asn Glu Ile
 35 40 45
- Gly Ile Lys Asn Asn Ile Ser Asn Phe Phe Ala Lys Ile Pro Phe Glu 50 55 60
- Glu Ala Asn Gly Arg Leu Val Ala Cys Thr Arg Thr Tyr Glu Glu Ser 65 70 75 80
- Ile Lys Gly Ser Cys Gly Gln Lys Glu Asn Lys Ile Lys Thr Val Ser 85 90 95
- Phe Glu Ser Lys Ile Gln Leu Arg Ser Lys Gln Glu Phe Gln Phe Phe 100 105 110
- Asp Glu Glu Glu Glu Thr Gly Glu Asn His Thr Ile Phe Ile Gly Pro 115 120 125
- Val Glu Lys Leu Ile Val Tyr Pro Pro Pro Pro Ala Lys Gly Gly Ile 130 135 140
- Ser Val Thr Asn Glu Asp Leu His Cys Leu Asn Glu Gly Glu Phe Leu 145 155 160
- Asn Asp Val Ile Ile Asp Phe Tyr Leu Lys Tyr Leu Val Leu Glu Lys 165 170 175
- Leu Lys Lys Glu Asp Ala Asp Arg Ile His Ile Phe Ser Ser Phe Phe
 180 185 190
- Tyr Lys Arg Leu Asn Gln Arg Glu Arg Arg Asn His Glu Thr Thr Asn 195 200 205
- Leu Ser Ile Gln Gln Lys Arg His Gly Arg Val Lys Thr Trp Thr Arg
- His Val Asp Ile Phe Glu Lys Asp Phe Ile Phe Val Pro Leu Asn Glu 225 235 240

-

Ala Ala His Trp Phe Leu Ala Val Cys Phe Pro Gly Leu Glu Lys 255 .

Pro Lys Tyr Glu 260 Pro Asn Pro His Tyr His Glu Asn Ala Val Ile Gln 270 Lys Cys Ser Thr Val Glu Asp Ser Cys Ile Ser Ser Ser Ala Ser Glu

Met Glu Ser Cys Ser Gln Asn Ser Ser Ala Lys Pro Val Ile Lys Lys 290 295 300

Met Leu Asn Lys Lys His Cys Ile Ala Val Ile Asp Ser Asn Pro Gly 305 310 315 320

Gln Glu Glu Ser Asp Pro Arg Tyr Lys Arg Asn Ile Cys Ser Val Lys 325 330 335

Tyr Ser Val Lys Lys Ile Asn His Thr Ala Ser Glu Asn Glu Glu Phe 340 345 350

Asn Lys Gly Glu Ser Thr Ser Gln Lys Ser 355

<210> 320

<211> 330

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (38)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (247)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 320

Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val 1 5 10 15

Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly 20 25 30

Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val 35 40 45

Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Asn Ile Trp Arg 50 55 60

Asp Leu Gly Met His Val Ile Ile Ala Gly Asn Asp Ser Lys Ala 65 70 75 80

Lys Gln Val Val Ser Lys Ile Lys Glu Glu Thr Leu Asn Asp Lys Val 85 90 95

1-11

- Glu Phe Leu Tyr Cys Asp Leu Ala Ser Met Thr Ser Ile Arg Gln Phe 100 105 110 110
- Val Gln Lys Phe Lys Met Lys Lys Ile Pro Leu His Val Leu Ile Asn 115 120 125
- Asn Ala Gly Val Met Met Val Pro Gln Arg Lys Thr Arg Asp Gly Phe 130 135 140
- Glu Glu His Phe Gly Leu Asn Tyr Leu Gly His Phe Leu Leu Thr Asn 145 150 155 160
- Leu Leu Leu Asp Thr Leu Lys Glu Ser Gly Ser Pro Gly His Ser Ala 165 . 170 . 175
- Arg Val Val Thr Val Ser Ser Ala Thr His Tyr Val Ala Glu Leu Asn 180 185 190
- Met Asp Asp Leu Gln Ser Ser Ala Cys Tyr Ser Pro His Ala Ala Tyr
 195 200 205
- Ala Gln Ser Lys Leu Ala Leu Val Leu Phe Thr Tyr His Leu Gln Arg
- Leu Leu Ala Ala Glu Gly Ser His Val Thr Ala Asn Val Val Asp Pro 235 235 240
- Gly Val Val Asn Thr Asp Xaa Tyr Lys His Val Phe Trp Ala Thr Arg 245 250 255
- Leu Ala Lys Lys Leu Leu Gly Trp Leu Leu Phe Lys Thr Pro Asp Glu 260 265 270
- Gly Ala Trp Thr Ser Ile Tyr Ala Ala Val Thr Pro Glu Leu Glu Gly 275 280 285
- Val Gly Gly Arg Tyr Leu Tyr Asn Glu Lys Glu Thr Lys Ser Leu His 290 295 300
- Val Thr Tyr Asn Gln Lys Leu Gln Gln Gln Leu Trp Ser Lys Ser Cys 305 310 315 320
- Glu Met Thr Gly Val Leu Asp Val Thr Leu 325 330
- <210> 321
- <211> 71
- <212> PRT
- <213> Homo sapiens
- <220>
- <221> SITE
- <222> (38)
- <223> Xaa equals any of the naturally occurring L-amino acids
- <400> 321
- Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val 1 5 10 15

<221> SITE <222> (174)

<220>
<221> SITE
<222> (195)

<220>
<221> SITE
<222> (199)

<220>
<221> SITE
<222> (206)

<400> 322

50

```
Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly
             20
                                  25
Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val
Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Asn Ile Trp Arg
     50
Asp Leu Ala Cys Met Leu Ser
<210> 322
<211> 266
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (97)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
```

<223> Xaa equals any of the naturally occurring L-amino acids

<223> Xaa equals any of the naturally occurring L-amino acids

<223> Xaa equals any of the naturally occurring L-amino acids

<223> Xaa equals any of the naturally occurring L-amino acids

Met Glu Val Thr Thr Glu Asp Thr Ser Arg Thr Asp Val Ser Glu Pro

Ala Thr Ser Gly Gly Ala Ala Asp Gly Val Thr Ser Ile Ala Pro Thr
20 25 30

Ala Val Ala Ser Ser Thr Thr Ala Ala Ser Ile Thr Thr Ala Ala Ser

Ser Met Thr Val Ala Ser Ser Ala Pro Thr Thr Ala Ala Ser Ser Thr

Thr Val Ala Ser Ile Ala Pro Thr Thr Thr Ala Ser Ser Met Thr Ala

75

-

Ala Ser Ser Thr Pro Met Thr Leu Ala Leu Pro Ala Pro Thr Ser Thr 85 90 95

Xaa Thr Gly Arg Thr Pro Ser Thr Thr Ala Thr Gly His Pro Ser Leu
100 105 110

Ser Thr Ala Leu Ala Gln Val Pro Lys Ser Ser Ala Leu Pro Arg Thr 115 120 125

Ala Thr Leu Ala Thr Leu Ala Thr Arg Ala Gln Thr Val Ala Thr Thr 130 140

Ala Asn Thr Ser Ser Pro Met Ser Thr Arg Pro Ser Pro Ser Lys His 155 150 155

Met Pro Ser Asp Thr Ala Ala Ser Pro Val Pro Pro Met Xaa Pro Gln
165 170 175

Ala Gln Gly Pro Ile Ser Gln Val Ser Val Asp Gln Pro Val Val Asn 180 185 190

Thr Thr Xaa Lys Ser Thr Xaa Met Pro Ser Asn Thr Thr Xaa Glu Pro 195 19

Leu Thr Gln Ala Val Val Asp Lys Thr Leu Leu Leu Val Val Leu Leu 210

Leu Gly Val Thr Leu Phe Ile Thr Val Leu Val Leu Phe Ala Leu Gln 235 240

Ala Tyr Glu Ser Tyr Lys Lys Lys Asp Tyr Thr Gln Val Asp Tyr Leu 245 250 255

Ile Asn Gly Met Tyr Ala Asp Ser Glu Met 260 265

<210> 323

<211> 99

<212> PRT

<213> Homo sapiens

<400> 323

Ala Arg Cys Pro Glu Leu Pro Gly Leu Arg Cys Arg Pro Arg Pro Arg

1 5 10 15

Gly Ala Cys Cys Ala Arg Met Arg Leu Leu Glu Trp Arg Val Tyr

Leu Arg Leu Thr Cys Ala Thr Lys Asp Gly Met Ala Arg Glu Cys Pro
50 55 60

Thr Thr Trp Leu Ser Pro Pro Ala Lys Pro Asp Phe Ala Gln Arg His 65 70 75 80

Ser Val Lys Pro Thr Ala Leu Gln Gly Gly Arg Trp Ser Arg Leu Gly 85 90 95

Ala Ser Pro

<210> 324

<211> 96

<212> PRT

<213> Homo sapiens

<400> 324

Leu Pro Ala Thr Val Glu Phe Ala Val His Thr Phe Asn Gln Gln Ser

1 10 15

Lys Asp Tyr Tyr Ala Tyr Arg Leu Gly His Ile Leu Asn Ser Trp Lys 20 25 30

Glu Gln Val Glu Ser Lys Thr Val Phe Ser Met Glu Leu Leu Gly 35 40

Arg Thr Arg Cys Gly Lys Phe Glu Asp Asp Ile Asp Asn Cys His Phe 50 60

Gln Glu Ser Thr Glu Leu Asn Asn Thr Phe Thr Cys Phe Phe Thr Ile 65 70 75 80

Ser Thr Arg Pro Trp Met Thr Gln Phe Ser Leu Leu Asn Lys Thr Cys 85 90 95

<210> 325

<211> 166

<212> PRT

<213> Homo sapiens

<400> 325

Leu Leu Trp Ala Arg Gly Leu Gly Arg Ala Lys Ser Ala Val Pro Thr 1 5 10

Val Ser Thr Met Leu Gly Leu Pro Trp Lys Gly Gly Leu Ser Trp Ala 20 25 30

Leu Leu Leu Leu Gly Ser Gln Ile Leu Leu Ile Tyr Ala Trp
35 40 45

His Phe His Glu Gln Arg Asp Cys Asp Glu His Asn Val Met Ala Arg 50 55

Tyr Leu Pro Ala Thr Val Glu Phe Ala Val His Thr Phe Asn Gln Gln 65 70 75 80

Ser Lys Asp Tyr Tyr Ala Tyr Arg Leu Gly His Ile Leu Asn Ser Trp 85 90 Lys Glu Gln Val Glu Ser Lys Thr Val Phe Ser Met Glu Leu Leu Leu 100 105 110 110

Gly Arg Thr Arg Cys Gly Lys Phe Glu Asp Asp Ile Asp Asn Cys His

Phe Gln Glu Ser Thr Glu Leu Asn Asn Thr Phe Thr Cys Phe Phe Thr 130 140

Ile Ser Thr Arg Pro Trp Met Thr Gln Phe Ser Leu Leu Asn Lys Thr 145 150 155 160

Cys Leu Glu Gly Phe His 165

<210> 326

<211> 214

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (200)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (205)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 326

Leu Glu Gln Lys Leu Glu Leu His Arg Gly Gly Gly Arg Ser Arg Thr 1 5 10 15

Ser Gly Ser Pro Gly Leu Gln Glu Phe Gly Thr Arg Glu Glu Arg Gly

Glu Gly Glu Gln Arg Thr Gly Arg Glu Phe Ser Gly Asn Gly Gly Arg
35 40 45

Ala Val Glu Ala Ala Arg Met Arg Leu Leu Cys Gly Leu Trp Leu Trp 50 55 60

Leu Ser Leu Leu Lys Val Leu Gln Ala Gln Thr Pro Thr Pro Leu Pro 65 70 75

Leu Pro Pro Pro Met Gln Ser Phe Gln Gly Asn Gln Phe Gln Gly Glu

Trp Phe Val Leu Gly Leu Ala Gly Asn Ser Phe Arg Pro Glu His Arg
100 105 110

Ala Leu Leu Asn Ala Phe Thr Ala Thr Phe Glu Leu Ser Asp Asp Gly 115 120 125

Arg Phe Glu Val Trp Asn Ala Met Thr Arg Gly Gln His Cys Asp Thr 130 140

Trp Ser Tyr Val Leu Ile Pro Ala Ala Gln Pro Gly Gln Phe Thr Val
145 155 160

Asp His Gly Val Gly Arg Ser Trp Leu Leu Pro Pro Gly Thr Leu Asp 165 170 175

Gln Phe Ile Cys Leu Gly Arg Ala Gln Gly Leu Ser Asp Asp Ile 180 185 190

Val Phe Pro Asp Val Thr Gly Xaa Ala Leu Asp Leu Xaa Ser Leu Pro 195 200 205

Trp Val Ala Ala Pro Ala 210

<210> 327

<211> 181

<212> PRT

<213> Homo sapiens

<400> 327

Met Cys Val Cys Glu Arg Lys Arg Gly Arg Glu Lys Glu Gly Gly Val

Thr Pro Thr Met Thr Ser Asn Phe Pro Phe Cys Thr Leu Ile Leu Gly
20 25 30

Ile Ala Gln Ala Gln Ala Cys Pro Gly Cys Pro Gly Asp Trp Pro Gly
35
40

Leu Gly Ser Gly Val Gly Glu Gly Leu His His Ile Arg Thr Cys Arg 50 55 60

Thr Pro Ile Pro Cys Ser Pro Pro Ala Pro Ala Ala Ala Cys Leu Gly 65 70 75 80

Ser Gly His Ala Arg Leu Pro Cys Val Leu Arg Leu Trp Pro Val Pro 85 90 95

Ala Asn Leu Ser Ser Pro Phe Arg Leu Glu Ala Leu His Cys Ser Phe 100 105 110

Trp Ser Ser Pro Leu Leu Pro Ala Pro His Leu Ala Phe Phe Gly Phe 115 120 125

Arg Asp Leu Leu Thr Asp Phe Leu Leu Ala Ala Cys Leu Leu Thr Phe 130 140

Gln Lys Thr Pro Leu Glu Leu Pro Met Ala Val Val His Leu Leu Val 145 150 155 160

Ala Thr Pro Cys Tyr Gln Met Leu Asp Asn Leu Pro Leu Pro Ser Ala 165 170 175

Ala Ala Asn Trp Cys 180 <210> 328

<211> 195

<212> PRT

<213> Homo sapiens

<400> 328

Tyr Leu Trp Gly Arg Pro Arg Leu Arg Met Arg Ala Gly Thr Ser Pro 1 5 10 15

Ser Ala Pro Trp Gly Glu Lys Arg Glu Lys Leu Gly His Lys Leu Pro 20 25 30

Val Ala Leu Gln Gly Tyr His Pro Trp Ile Leu Leu Glu Cys Thr Val 35 40 45

Phe Trp Ala Arg Val Val Leu Ala Cys Phe Ser Leu Tyr Leu Ile Arg 50 55 60

Gly Pro Asn Cys Ile Asn Arg Gln Pro Glu Pro Thr Tyr Gln Lys Ala
65 70 75 80

Cys Asn Leu Asp Cys Ser Ser Asp Phe Gly Gln Glu Arg Ala Pro Ala 85 90 95

Trp Glu Leu Leu Gly Pro Glu Ser Glu Gln Arg Leu Arg Glu Tyr Thr 100 105 110

Ala Gln Gly Leu Gln Ser Leu Ala Ser Ser His Arg Trp Arg Gln Phe 115 120 125

Lys Thr Glu Gly Lys Met Arg Gly Gly Ala Ser Pro Leu Pro Trp Leu 130 $$ 135 $$ 140

Ile Cys Phe Trp Leu Cys Ser Tyr Lys Gly Ser Asp Asn Ser Leu Lys 145 150 155 160

Pro Val Val Pro Gly Pro Thr Leu Cys Pro Gln Ser Leu Val Ser Pro 165 170 175

Ser Val His Pro Ser Thr Arg Ser Ala Ser Leu Gly Arg His Arg Ala 180 185 190

Glu Ala Ala 195

<210> 329

<211> 50

<212> PRT

<213> Homo sapiens

<400> 329

Met Pro Gly Ile Leu Ala Gly Ile Pro Val Lys Asp Leu Cys Leu Ser 1 5 10 15

Leu Leu Gln Gly Phe Arg Leu Leu Leu Cys Val Cys Pro Gly Trp
20 25 30

Leu Ser Gly Trp Met Gly Gly Gln Lys Gly Ser Pro Arg Ile Val Asp

35 40 45

Ile Gly 50

<210> 330

<211> 90

<212> PRT.

<213> Homo sapiens

<400> 330

Ala Lys Gly Glu Glu Arg Lys Glu Ala Phe Ser Leu Lys Met Val Gln

Leu Ser Ser Glu Pro Ile Ser Phe Gly Leu Met Tyr Leu Tyr Leu Gly 25 . 20

Val Phe Phe His Leu Ile Tyr Pro Gly Ala Leu Ser Ile Thr Thr Leu

Gly Lys His Ser His Pro Phe Phe Thr Ala Glu Gln Asn Ser Thr Val

Trp Met Glu His Thr Leu Phe His Gln Ser Pro Val Ala Ser His Leu 75 65 70

Val Cys Phe Gln Ser Phe Ala Phe Ser Glu

<210> 331

<211> 56

<212> PRT

<213> Homo sapiens

<400> 331

Gly Pro Ala His Pro Ala Ser Pro Pro Leu Met Thr Leu Ser Leu Gln

Leu Ala Glu Leu Val His Phe Val Cys Ala Phe Gln Ser Gln Trp Thr

Gly Val Tyr Pro Met Met Pro Pro Leu Lys Pro Thr Glu Pro Leu Cys 40

Phe Ala Cys Val Pro Cys Arg Val 50

<210> 332

<211> 18

<212> PRT

<213> Homo sapiens

<400> 332

Met Leu Leu Glu Val Tyr Gly Asp Ser Ile Ser Val Thr Val Ala Ile 10

```
Pro Leu
```

<210> 333
<211> 19
<212> PRT
<213> Homo sapiens
<400> 333
Met His Ser Pro Cys Gln Ser Lys Ala Ala Asp Gly Leu Gly Lys Ser

Glu Thr Glu

<210> 334 <211> 10 <212> PRT <213> Homo sapiens <400> 334

Met Leu Lys Ser Leu Gly Leu Ser Thr Asn 1 5 10

<210> 335 <211> 200 <212> PRT <213> Homo sapiens

<400> 335
Ala Gln Arg Leu Ala Glu Glu Cys Phe Tyr Met Leu Leu Glu Val Tyr
1 5 10 15

Gly Asp Ser Ile Ser Val Thr Val Ala Ile Pro Leu Met His Ser Pro 20 25 30

Cys Gln Ser Lys Ala Ala Asp Gly Leu Gly Lys Ser Glu Thr Glu Met 35 40 45

Leu Lys Ser Leu Gly Leu Ser Thr Asn Met Ser Pro Phe His Leu Leu 50 60

Gly Leu Lys Val Phe Leu Thr Trp Ala Leu Thr Leu Ala Gln Ile Cys 65 70 75 80

Leu Tyr Phe Phe Glu Val Gln Pro Leu Gly Leu Leu Ala Leu Asn Phe 85 90 95

Phe Cys Thr Ala Thr Ala Gly Leu Lys Glu Leu Cys Met His Pro Pro 100 105 110

Ser Leu Ala Phe Thr Pro Glu Phe His Thr Ser Leu Ser Pro Leu Ala 115 120 125

Ile Pro Ser Phe Cys Gly Thr Ser Val Ser Leu Ser Asn Ser His Thr 130 135 140

Ile Pro Leu Ser Leu Tyr Leu Pro Phe Pro Ser Lys Ser Arg Met Pro 150 155

Asp Thr Leu His Leu Leu Val His Ser Leu Pro Leu Val His Ser Gln 165

Val Leu Pro Val Lys Asp Val Thr Ile Glu Trp Pro Leu Cys Gln Arg 180 185

Cys Leu Gly Ser Thr Cys His Gln 195

<210> 336

<211> 99

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (94)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (99)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 336

Trp Ile Pro Arg Ala Ala Gly Ile Arg His Glu Val Gln Val Ser Leu

Phe Gln Met Phe Cys Phe Ser Ser Ile Phe Cys Ser His Glu His Thr

His Leu Pro Gly Thr Phe Trp Leu Phe Leu Phe Leu Phe Leu Ile Leu

Pro Pro Ser Cys Pro Cys Phe Leu Pro Phe Ser Leu Ala Ile Glu Thr 55 60

Val Arg Trp Pro Cys Trp His His Pro Thr Ser Phe Glu Leu Cys Tyr 70

Pro Gly Thr Ser Ile Tyr Tyr Ala Ser Arg Gly Gly Pro Xaa Pro Asn

Ser Glu Xaa

<210> 337

<211> 96

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

Ser Lys Val Pro Arg Leu Leu Ile Ile Gln Cys Pro Gln Gln Arg Glu 85 90 95

Pro Val Ser Tyr Trp Leu
50

<210> 339 <211> 287 <212> PRT <213> Homo sapiens <400> 339

Pro Arg Val Arg Lys Glu Pro Glu Ala Met Gln Trp Leu Arg Val Arg
1 5 10 15

- Glu Ser Pro Gly Glu Ala Thr Gly His Arg Val Thr Met Gly Thr Ala 20 25 30
- Ala Leu Gly Pro Val Trp Ala Ala Leu Leu Leu Phe Leu Leu Met Cys 35
- Glu Ile Pro Met Val Glu Leu Thr Phe Asp Arg Ala Val Ala Ser Asp 50 60
- Cys Gln Arg Cys Cys Asp Ser Glu Asp Pro Leu Asp Pro Ala His Val 65 70 75 80
- Ser Ser Ala Ser Ser Ser Gly Arg Pro His Ala Leu Pro Glu Ile Arg 85 . 90 95
- Pro Tyr Ile Asn Ile Thr Ile Leu Lys Gly Asp Lys Gly Asp Pro Gly 100 105 110
- Pro Met Gly Leu Pro Gly Tyr Met Gly Arg Glu Gly Pro Gln Gly Glu 115 120 125
- Pro Gly Pro Gln Gly Ser Lys Gly Asp Lys Gly Glu Met Gly Ser Pro 130 135 140
- Gly Ala Pro Cys Gln Lys Arg Phe Phe Ala Phe Ser Val Gly Arg Lys 145 150 155 160
- Thr Ala Leu His Ser Gly Glu Asp Phe Gln Thr Leu Leu Phe Glu Arg 165 170 175
- Val Phe Val Asn Leu Asp Gly Cys Phe Asp Met Ala Thr Gly Gln Phe 180 185 190
- Ala Ala Pro Leu Arg Gly Ile Tyr Phe Phe Ser Leu Asn Val His Ser 195 200 205
- Trp Asn Tyr Lys Glu Thr Tyr Val His Ile Met His Asn Gln Lys Glu 210 215 220
- Ala Val Ile Leu Tyr Ala Gln Pro Ser Glu Arg Ser Ile Met Gln Ser 225 230 235 240
- Gln Ser Val Met Leu Asp Leu Ala Tyr Gly Asp Arg Val Trp Val Arg 245 250 255
- Leu Phe Lys Arg Gln Arg Glu Asn Ala Ile Tyr Ser Asn Asp Phe Asp 260 265 270
- Thr Tyr Ile Thr Phe Ser Gly His Leu Ile Lys Ala Glu Asp Asp 275 280 285
- <210> 340
- <211> 339
- <212> PRT
- <213> Homo sapiens
- <400> 340
- Met Leu Tyr Pro Gly Ser Val Tyr Leu Leu Gln Lys Ala Leu Met Pro

	1				5				1	.0				1	.5
Va.	l Le	u Le	eu Gl 2	n Gl;	y Gl	n Ala	a Arg	g Le		l Gl	u Gl	u Cy	s As 3		y Ar
Ar	g Al	а Ly 3	s Le 5	u Lei	u Ala	a Cys	Asp 40		y As	n Gl	u Il	e Ası		r Me	t Ph
Va.	l As 5	p Ar 0	g Ar	g Gly	/ Thi	Ala 55	Glu	ı Pro	o Gla	n Gly	y Gli 60		s Lei	ı Va	1 11
Суз 65	s Cy	s Gl	u Gl	y Asr	n Ala 70	a Gly	Phe	туз	c Glu	u Va] 79		y Cys	s Val	l Se	r Th
Pro	Le	u Gl	u Al	a Gly 85	r Tyr	Ser	Val	Leu	Gl ₃ 90) Asr	n His	Pro	Gl ₃	
Ala	Gly	y Se	r Th:	r Gly	Val	Pro	Phe	Pro 105	Glr	n Asn	ı Glu	ı Ala	110		a Męt
Asp	Va]	l Vaj	l Val	l Gln	Phe	Ala	Ile 120	His	Arg	, Leu	Gly	Phe 125		. Pro	Glr
Asp	11e	e Ile	≘ Ile	Yyr	Ala	Trp 135	Ser	Ile	Gly	Gly	Phe 140		Ala	Thr	Trp
Ala 145	Ala	Met	. Ser	Tyr	Pro 150	Asp	Val	Ser	Ala	Met 155	Ile	Leu	Asp	Ala	Ser 160
				Val 165					170			•		175	
Arg	Gly	Leu	Val 180	Thr	Arg	Thr	Val	Arg 185	Gln	His	Leu		Leu 190	Asn	Asn
Ala	Glu	Gln 195	Leu	Cys	Arg	Tyr	Gln 200	Gly	Pro	Val	Leu	Leu 205	Ile	Arg	Arg
Thr	Lys 210	Asp	Glu	Ile	Ile	Thr 215	Thr	Thr	Val	Pro	Glu 220	Asp	Ile	Met	Ser
Asn 225	Arg	Gly	Asn	Asp	Leu 230	Leu	Leu	Lys	Leu	Leu 235	Gln	His	Arg	Tyr	Pro 240
Arg	Val	Met	Ala	Glu 245	Glu	Gly :	Leu .	Arg	Val 250	Val	Arg	Gln	Trp	Leu 255	Glu
Ala	Ser	Ser	Gln 260	Leu	Glu	Glu A	Ala :	Ser 265	Ile	Tyr	Ser	Arg	Trp 270	Glu	Val
Glu	Glu	Asp 275	Trp	Cys	Leu	Ser 1	Val 1 280	Leu	Arg	Ser		Gln 285	Ala	Glu	His
Gly	Pro 290	Asp	Phe	Pro	Trp	Ser 1 295	/al (Gly	Glu		Met 300	Ser .	Ala .	Asp	Gly
Arg 305	Arg	Gln	Leu	Ala	Leu 1 310	Phe I	eu A	Ala .	Arg	Lys :	His :	Leu 1	His .		Phe 320

WO 00/06698 PCT/US99/17130

227

Glu Ala Thr His Cys Thr Pro Leu Pro Ala Gln Asn Phe Gln Met Pro 325 330335

Trp His Leu

<210> 341

<211> 127

<212> PRT

<213> Homo sapiens

<400> 341

Val Cys Pro Lys Trp Cys Arg Phe Leu Thr Met Leu Gly His Cys Cys

1 10 15

Tyr Phe Trp Gln Val Trp Pro Ala Ser Glu Ala Leu Ala Ala Gly Pro 20 25 30

Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser Trp Lys Gln His Ile Gly 35 40 45

Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu Pro Thr Thr Thr Leu Thr 50 60

Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly Lys Asn Pro Ala Ala Gly 65 70 75 80

Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly Val Trp Pro Cys Phe Ala 85 90 95

Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr Pro Ser Ser Thr Gly Leu 100 105 110

Arg Ser Cys Leu Val Arg Ser Pro Ala Thr Trp Trp Arg Thr Pro 115 120 125

<210> 342

<211> 554

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (16)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (109)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 342

Trp Ile Pro Arg Ala Ala Gly Ile Arg His Glu Ile Tyr Arg Glu Xaa 1 5 10 15

Asp Ser Glu Arg Ala Pro Ala Ser Val Pro Glu Thr Pro Thr Ala Val 20 25 30

- Thr Ala Pro His Ser Ser Ser Trp Asp Thr Tyr Tyr Gln Pro Arg Ala 35 40 45
- Leu Glu Lys His Ala Asp Ser Ile Leu Ala Leu Ala Ser Val Phe Trp 50 55 60
- Ser Ile Ser Tyr Tyr Ser Ser Pro Phe Ala Phe Phe Tyr Leu Tyr Arg 65 70 75 80
- Lys Gly Tyr Leu Ser Leu Ser Lys Val Val Pro Phe Ser His Tyr Ala 85 90 95
- Gly Thr Leu Leu Leu Leu Leu Ala Gly Val Ala Cys Xaa Arg Gly Ile 100 105 110
- Gly Arg Trp Thr Asn Pro Gln Tyr Arg Gln Phe Ile Thr Ile Leu Glu 115 120 125
- Ala Thr His Arg Asn Gln Ser Ser Glu Asn Lys Arg Gln Leu Ala Asn 130 135 140
- Tyr Asn Phe Asp Phe Arg Ser Trp Pro Val Asp Phe His Trp Glu Glu 145 150 155 160
- Pro Ser Ser Arg Lys Glu Ser Arg Gly Gly Pro Ser Arg Arg Gly Val 165 170 175
- Ala Leu Leu Arg Pro Glu Pro Leu His Arg Gly Thr Ala Asp Thr Leu 180 185 190
- Leu Asn Arg Val Lys Lys Leu Pro Cys Gln Ile Thr Ser Tyr Leu Val
- Ala His Thr Leu Gly Arg Arg Met Leu Tyr Pro Gly Ser Val Tyr Leu 210 215 220
- Leu Gln Lys Ala Leu Met Pro Val Leu Leu Gln Gly Gln Ala Arg Leu 225 230 230 240
- Val Glu Glu Cys Asn Gly Arg Arg Ala Lys Leu Leu Ala Cys Asp Gly 245 250 250
- Asn Glu Ile Asp Thr Met Phe Val Asp Arg Arg Gly Thr Ala Glu Pro 260 265 270
- Gln Gly Gln Lys Leu Val Ile Cys Cys Glu Gly Asn Ala Gly Phe Tyr 275 280 285
- Glu Val Gly Cys Val Ser Thr Pro Leu Glu Ala Gly Tyr Ser Val Leu 290 295 300
- Gly Trp Asn His Pro Gly Phe Ala Gly Ser Thr Gly Val Pro Phe Pro 305 310 315 320
- Arg Leu Gly Phe Gln Pro Gln Asp Ile Ile Ile Tyr Ala Trp Ser Ile

340 345 350 Gly Gly Phe Thr Ala Thr Trp Ala Ala Met Ser Tyr Pro Asp Val Ser Ala Met Ile Leu Asp Ala Ser Phe Asp Asp Leu Val Pro Leu Ala Leu 375 Lys Val Met Pro Asp Ser Trp Arg Gly Leu Val Thr Arg Thr Val Arg 395 Gln His Leu Asn Leu Asn Asn Ala Glu Gln Leu Cys Arg Tyr Gln Gly 405 Pro Val Leu Leu Ile Arg Arg Thr Lys Asp Glu Ile Ile Thr Thr Thr 425 Val Pro Glu Asp Ile Met Ser Asn Arg Gly Asn Asp Leu Leu Lys 440 Leu Leu Gln His Arg Tyr Pro Arg Val Met Ala Glu Glu Gly Leu Arg 455 Val Val Arg Gln Trp Leu Glu Ala Ser Ser Gln Leu Glu Glu Ala Ser Ile Tyr Ser Arg Trp Glu Val Glu Glu Asp Trp Cys Leu Ser Val Leu . 485 490 Arg Ser Tyr Gln Ala Glu His Gly Pro Asp Phe Pro Trp Ser Val Gly Glu Asp Met Ser Ala Asp Gly Arg Arg Gln Leu Ala Leu Phe Leu Ala Arg Lys His Leu His Asn Phe Glu Ala Thr His Cys Thr Pro Leu Pro 540 Ala Gln Asn Phe Gln Met Pro Trp His Leu <210> 343 <211> 225 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (5) <223> Xaa equals any of the naturally occurring L-amino acids <400> 343 His Glu Arg Ala Xaa Gly Pro Ser Arg Gly His Gly Glu Leu Leu Ser

Cys Val Leu Gly Pro Arg Leu Tyr Lys Ile Tyr Arg Glu Arg Asp Ser

20

Glu Arg Ala Pro Ala Ser Val Pro Glu Thr Pro Thr Ala Val Thr Ala 35 40 45

Pro His Ser Ser Ser Trp Asp Thr Tyr Tyr Gln Pro Arg Ala Leu Glu
50 55 60

Lys His Ala Asp Ser Ile Leu Ala Leu Ala Ser Val Phe Trp Ser Ile 65 70 75 80

Ser Tyr Tyr Ser Ser Pro Phe Ala Phe Phe Tyr Leu Tyr Arg Lys Gly 85 90 95

Tyr Leu Ser Leu Ser Lys Val Val Pro Phe Ser His Tyr Ala Gly Thr 100 105. 110

Leu Leu Leu Leu Ala Gly Val Ala Cys Ser Glu Ala Leu Ala Ala 115 120 125

Gly Pro Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser Trp Lys Gln His

Ile Gly Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu Pro Thr Thr Thr 145 150 155 160

Leu Thr Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly Lys Asn Pro Ala 165 170 175

Ala Gly Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly Val Trp Pro Cys
180 185 190

Phe Ala Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr Pro Ser Ser Thr 195 200 205

Gly Leu Arg Ser Cys Leu Val Arg Ser Pro Ala Thr Trp Trp Arg Thr 210 225 220

Pro 225

<210> 344

<211> 299

<212> PRT

<213> Homo sapiens

<400> 344

Met Phe Lys Arg His Gln Arg Leu Lys Lys Asp Ser Thr Gln Ala Glu

1 10 15

Glu Asp Leu Ser Glu Gln Glu Gln Asn Gln Leu Asn Val Leu Lys Lys 20 25 30

His Gly Tyr Val Val Gly Arg Val Gly Arg Thr Phe Leu Tyr Ser Glu 35 . 40 45

Glu Gln Lys Asp Asn Ile Pro Phe Glu Phe Asp Ala Asp Ser Leu Ala 50 55 60

Phe Asp Met Glu Asn Asp Pro Val Met Gly Thr His Lys Ser Thr Lys

65 75 70 Gln Val Glu Leu Thr Ala Gln Asp Val Lys Asp Ala His Trp Phe Tyr Asp Thr Pro Gly Ile Thr Lys Glu Asn Cys Ile Leu Asn Leu Leu Thr Glu Lys Glu Val Asn Ile Val Leu Pro Thr Gln Ser Ile Val Pro Arg 120 Thr Phe Val Leu Lys Pro Gly Met Val Leu Phe Leu Gly Ala Ile Gly 135 Arg Ile Asp Phe Leu Gln Gly Asn Gln Ser Ala Trp Phe Thr Val Val 155 Ala Ser Asn Ile Leu Pro Val His Ile Thr Ser Leu Asp Arg Ala Asp 165 170 Ala Leu Tyr Gln Lys His Ala Gly His Thr Leu Leu Gln Ile Pro Met Gly Gly Lys Glu Arg Met Ala Gly Phe Pro Pro Leu Val Ala Glu Asp 200 Ile Met Leu Lys Glu Gly Leu Gly Ala Ser Glu Ala Val Ala Asp Ile Lys Phe Ser Ser Ala Gly Trp Val Ser Val Thr Pro Asn Phe Lys Asp Arg Leu His Leu Arg Gly Tyr Thr Pro Glu Gly Thr Val Leu Thr Val Arg Pro Pro Leu Leu Pro Tyr Ile Val Asn Ile Lys Gly Gln Arg Ile 265 Lys Lys Ser Val Ala Tyr Lys Thr Lys Lys Pro Pro Ser Leu Met Tyr 280 Asn Val Arg Lys Lys Gly Lys Ile Asn Val 290 295 <210> 345 <211> 314 <212> PRT <213> Homo sapiens <220> <221> SITE <222> (147) <223> Xaa equals any of the naturally occurring L-amino acids <220> <221> SITE <222> (211) <223> Xaa equals any of the naturally occurring L-amino acids

Pro Glu Pro Gln Pro Bro Bro Bro Bro Bro Bro Bro Bro Bro B				!> :																	•
20 25 30 Glu Arg Arg Cys Ala Ala Ala Ala Ser Ser Phe Gln His Ser Ser Ser Let 40 40 85 87 86 87 88 88 88 88 88 88 88 88 88 88 88 88		Me	t 1	Leu	ı Pr	O A.	la A	rg L 5	eu P	ro	Ph∈	Ar			Let	ı Se	r Le	eu F	he		
Gly Arg Glu Leu Pro Tyr Asp Pro Val Asp Thr Glu Gly Phe Gly Gly Gly Asp Met Gln Glu Arg Phe Leu Phe Pro Glu Tyr Ile Leu As 75 Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Glu Glu Glu Glu Glu Glu Arg Gln Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Arg Arg Glu Glu Arg Ile Arg Gln Gln Asp Arg Gln Gln Asp Arg Glu Glu Arg Ile Arg Inc Arg Ile Arg Ile Arg Inc Arg Ile Arg Ile Arg Inc Arg Inc Arg Ile Arg Ile Arg Ile Arg Inc Arg Ile Arg Ile Arg Inc Arg Inc Arg Inc Arg Ile Arg Ile Arg Inc Arg Inc Arg Inc Arg Inc Arg Ile Arg Inc Arg Inc Arg Inc Arg Inc Arg Ile Arg Inc		Gl	У	Ser	Al	a Pi	го Тì 20	nr A	la A	la	Arg			ly	Let	ı Ar	g G]	lu F		Let	ı Le
Signature of the state of the s		Gl	u.	Arg	Ar 3	g C <u>y</u> 5	/s A]	la A	la A	la	Ser 40	Se	r Pl	he	Glr	ı Hi			er	Ser	Le
Pro Glu Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Glu Arg Gln Gln Gln Glu Glu Glu Glu Glu Arg Gln Arg Gln Gln Arg Arg Glu Glu Arg Gln Gln Arg Gln Arg Glu His Pro Val Val Glu 115		Gl	У	Arg 50	G1	u Le	u Pr	:0 Ту			Pro	Va:	l As	sp	Thr			у Р	he	Gly	Gl
Gln Gln Gln Glu Glu Glu Glu Rrg Gln Arg Gln Gln Arg Glu Glu Arg Gln Gln Gln Arg Gln Gln Gln Tyr Leu Gln Leu Gln Arg Arg Arg Ala Leu 185 Arg Leu Gln Val Ser Arg Gln Gln Tyr Leu Gln Leu Val Ser Ala Ala 205 Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Arg Leu Leu 215 Arg Leu Pro Arg Ala Leu Leu Pro Arg Leu Pro Arg Leu Val Gly Pro 225 Ala Pro Gly Tyr Arg Gln Arg Leu Arg Gln Arg Leu Trp Gln Arg 255 Ala Arg Ala Gly Leu Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 275 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Gl ₃	y (Gly	As	p Me	t Gl	n Gl	lu Ai 70	rg	Phe	Let	ı Pl	ne			ту ту	r I	le	Leu	Ası 80
Arg Gln Gln Asn Leu Arg Ala Arg Ser Arg Glu His Pro Val Val Gl 115		Pro	o (3lu	Pro	o Gl	n Pr 8	o Th 5	ır Aı	g.	Glu	Lys			Leu	Glr	ı Gl	u L	eu		
His Pro Asp Pro Ala Leu Pro Pro Ser Gly Val Asn Cys Ser Gly Cyn 130 Gly Ala Xaa Leu His Cys Gln Asp Ala Gly Val Pro Gly Tyr Leu Pro 155 Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg 165 Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu 195 Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala 195 Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu 210 Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 235 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Ala Pro Gly Thr Lys Gly His Ser Ala Ala 275 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Glr	1 (3ln	Glı	1 Gl	u Gl 0	u Gl	u Ar	g	Gln	Arg 105	r Gl	.n	Gln	Arg	Ar			Glu	Arg
Gly Ala Xaa Leu His Cys Gln Asp Ala Gly Val Pro Gly Tyr Leu Pro 150 Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg 175 Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu 180 Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala 205 Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu 215 Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 235 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Ala Pro Gly Thr Lys Gly His Ser Ala 280 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Arg	j G	Sln	Glr 115	ı Ası	n Le	u Ar	g Al	.a .	Arg 120	Ser	Ar	g	Glu	His			al	Val	Gly
Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg Thr Val Cys Gln Arg Cys Trp Leu Leu Leu Ser His His Arg Arg Ala Leu 195		His	: F	ro .30	Asp	Pro	o Ala	a Le	u Pr 13	o 1	Pro	Ser	Gl	У	Val			s Se	er	Gly	Cys
Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu 185 Leu Arg Lau Pro Asp Ala Leu 230 Lys Gln Leu Ile Val Leu Gly Asn Lys Ser Leu Pro Gly Asp Cys 245 Ala Pro Gly Tyr Arg Gln Arg Leu Ala Pro Gly Thr Lys Gly His Ser Ala Ala 286 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Gly 145	Α	la	Xaa	Lei	ı His	з Су 15	s Gl O	n A	Asp	Ala	Gl			Pro	Gly	/ Ту	r:	Leu	Pro 160
Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala 205 Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu 210 Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 225 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Arg Glu Arg Leu Trp Glu Asp Cys 260 Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Tyr Lys Gly His Ser Ala 280 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Arg	G	lu	Lys	Phe	Le: 165	ı Ar	g Th	r A	Ala	Glu			Asp	Gly	Gly	'Le			Arg
Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu Asp Leu Leu 210 Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 240 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Arg Glu Arg Leu Trp Glu Asp Cys 260 Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 280 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr	ı	Thr	V	al	Cys	Gln 180	a Arg	ј Су:	s Tr	рI			Se	r I	lis	His	Arg			Ala	Leu
Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro 240 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Arg 265 Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 280 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr		Arg	L	eu	Gln 195	Val	. Ser	Arg	g Gl	u G 2	31n 200	Tyr	Let	ı (Slu	Leu			r A	Ala	Ala
Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp 255 Ala Pro Gly Tyr Arg Gln Arg Leu Arg 265 Glu Arg Leu Trp Glu Asp Cys 260 Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 275 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr	3	Leu	A: 2:	rg :	Xaa	Pro	Gly	Pro	Se:	r L	eu ¹	Val	Leu	1 T			Val	As	ρI	eu	Leu
Ala Pro Gly Tyr Arg Gln Arg Leu Arg 265 Glu Arg Leu Trp Glu Asp Cys 260 Gly Arg Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 275 Thr Gly Arg Ile Arg Ile Arg Arg Thr	2	Asp 225	Le	≘u :	Pro	Asp	Ala	Let 230	Leu)	ı P	ro i	Asp	Leu			Ala	Leu	Va.	1 G		Pro 240
Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala 275 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Thr	I	ys	G]	ln 1	Leu	Ile	Val 245	Leu	Gly	7 A	sn I	Lys	Val 250	. A	sp 1	Leu	Leu	Pro			Asp
Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Thr	P	la	Pr	:o (Sly	Туг 260	Arg	Gln	Arg	L	eu A	arg 265	Glu	. A	rg l	Leu	Trp			sp (Cys
	A	la	Ar	g 2	Ala 275	Gly	Leu	Leu	Leu	A. 28	la F 80	ro	Gly	T	hr I	ys		His	S	er i	Ala
	P	ro	Se 29	r A	arg	Thr	Ser	His	Arg 295	Tì	hr G	ly.	Arg	I			Ile	Arg	A :	rg 1	Thr

Gly Pro Ala Gln Trp Ser Gly Thr Cys Gly 305 310

<210> 346

<211> 380

<212> PRT

<213> Homo sapiens

<400> 346

Pro Ser Phe Arg Arg Glu Arg Val Glu Thr Gly Gly Gly Pro Val
1 5 10 15

Thr His Gly Thr Glu Gly Pro Phe Leu Pro Leu Pro Gly Gly Thr Arg

Met Asn Met Thr Gln Ala Arg Val Leu Val Ala Ala Val Val Gly Leu 35 40 45

Val Ala Val Leu Leu Tyr Ala Ser Ile His Lys Ile Glu Glu Gly His 50 55 60

Leu Ala Val Tyr Tyr Arg Gly Gly Ala Leu Leu Thr Ser Pro Ser Gly 65 70 75 80

Pro Gly Tyr His Ile Met Leu Pro Phe Ile Thr Thr Phe Arg Ser Val 85 90 95

Gln Thr Thr Leu Gln Thr Asp Glu Val Lys Asn Val Pro Cys Gly Thr
100 105 110

Ser Gly Gly Val Met Ile Tyr Ile Asp Arg Ile Glu Val Val Asn Met 115 120 125

Leu Ala Pro Tyr Ala Val Phe Asp Ile Val Arg Asn Tyr Thr Ala Asp 130 135 140

Tyr Asp Lys Thr Leu Ile Phe Asn Lys Ile His His Glu Leu Asn Gln 145 150 155 160

Phe Cys Ser Ala His Thr Leu Gln Glu Val Tyr Ile Glu Leu Phe Asp 165 170 175

Gln Ile Asp Glu Asn Leu Lys Gln Ala Leu Gln Lys Asp Leu Asn Leu 180 185 190

Met Ala Pro Gly Leu Thr Ile Gln Ala Val Arg Val Thr Lys Pro Lys 195 200 205

Ile Pro Glu Ala Ile Arg Arg Asn Phe Glu Leu Met Glu Ala Glu Lys 210 215 220

Thr Lys Leu Leu Ile Ala Ala Gln Lys Gln Lys Val Val Glu Lys Glu 225 230 235 240

Ala Glu Thr Glu Arg Lys Lys Ala Val Ile Glu Ala Glu Lys Ile Ala 245 250 255

Gln Val Ala Lys Ile Arg Phe Gln Gln Lys Val Met Glu Lys Glu Thr

260 265 270

Glu Lys Arg Ile Ser Glu Ile Glu Asp Ala Ala Phe Leu Ala Arg Glu 275 280 285

Lys Ala Lys Ala Asp Ala Glu Tyr Tyr Ala Ala His Lys Tyr Ala Thr 290 295 300

Ser Asn Lys His Lys Leu Thr Pro Glu Tyr Leu Glu Leu Lys Lys Tyr 305 310 315 320

Gln Ala Ile Ala Ser Asn Ser Lys Ile Tyr Phe Gly Ser Asn Ile Pro 325 330 335

Asn Met Phe Val Asp Ser Ser Cys Ala Leu Lys Tyr Ser Asp Ile Arg 340 345 350

Thr Gly Arg Glu Ser Ser Leu Pro Ser Lys Glu Ala Leu Glu Pro Ser 355 360 365

Gly Glu Asn Val Ile Gln Asn Lys Glu Ser Thr Gly 370 375 380

<210> 347

<211> 422

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (328)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 347

Trp Ser Thr Gly Asn Ala Ser Trp Glu Lys Lys Asp Asn Phe Ile Leu 1 5 10 15

Ser Ala Asp Phe Glu Met Met Gly Leu Gly Asn Gly Arg Arg Ser Met 20 25 30

Lys Ser Pro Pro Leu Val Leu Ala Ala Leu Val Ala Cys Ile Ile Val 35 40 45

Leu Gly Phe Asn Tyr Trp Ile Ala Ser Ser Arg Ser Val Asp Leu Gln 50 55 60

Thr Arg Ile Met Glu Leu Glu Gly Arg Val Arg Arg Arg Ala Ala Glu 65 70 75 80

Arg Gly Ala Val Glu Leu Lys Lys Asn Glu Phe Gln Gly Glu Leu Glu 85 90 95

Lys Gln Arg Glu Gln Leu Asp Lys Ile Gln Ser Ser His Asn Phe Gln
100 105 110

Leu Glu Ser Val Asn Lys Leu Tyr Gln Asp Glu Lys Ala Val Leu Val 115 120 125

Asn	Asn 130		Thr	Thr	Gly	Glu 135		, Let	ı Ile	e Arg	7 Val 140		ı Glr	n Asp	Gl:
Leu 145		Thr	Leu	Gln	Arg 150		Tyr	Gly	/ Arg	155		Glr	n Asp	Va]	Let 160
Gln	Phe	Gln	Lys	Asn 165		Thr	Asn	Leu	170		Lys	Phe	Ser	Туг 175	
Leu	Ser	Gln	Cys 180	Ile	Asn	Gln	Met	Lys 185		val	Lys	Glu	190		Glu
Glu	Arg	Ile 195		Glu	Val	Thr	Lys 200		Gly	Asn	Glu	Ala 205		Ala	Sei
Arg	Asp 210	Leu	Ser	Glu	Asn	Asn 215		Gln	Arg	Gln	Gln 220		Gln	Ala	Leu
Ser 225	Glu	Pro	Gln	Pro	Arg 230	Leu	Gln	Ala	Ala	Gly 235	Leu	Pro	His	Thr	Glu 240
Val	Pro	Gln	Gly	Lys 245	Gly	Asn	Val	Leu	Gly 250	Asn	Ser	Lys	Ser	Gln 255	
Pro	Ala	Pro	Ser 260	Ser	Glu	Val	Val	Leu 265	Asp	Ser	Lys	Arg	Gln 270	Val	Glu
Lys	_, Glu	Glu 275	Thr	Asn	Glu	Ile	Gln 280	Val	Val	Asn	Glu	Glu 285	Pro	Gln	Arg
Asp	Arg 290	Leu	Pro	Gln	Glu	Pro 295	Gly	Arg	Glu	Gln	Val 300	Val	Glu	Asp	Arg
Pro 305	Val	Gly	Gly	Arg	Gly 310	Phe	Gly	Gly	Ala	Gly 315	Glu	Leu	Gly	Gln	Thr 320
Pro	Gln	Val	Gln	Ala 325	Ala	Leu	Xaa	Val	Ser 330	Gln	Glu	Asn	Pro	Glu 335	Met
Glu	Gly	Pro	Glu 340	Arg	qaA	Gln	Leu	Val 345	Ile	Pro	Asp	Gly	Gln 350	Glu	Glu
Glu	Gln	Glu 355	Ala	Ala	Gly	Glu	Gly 360	Arg	Asn	Gln	Gln	Lys 365	Leu	Arg	Gly
3lu	Asp 370	Asp	Tyr	Asn	Met	Asp 375	Glu	Asn	Glu	Ala	Glu 380	Ser	Glu	Thr	Asp
Lys 385	Gln	Ala	Ala	Leu	Ala 390	Gly	Asn	Asp	Arg	Asn 395	Ile	Asp	Val	Phe	Asn 400
/al	Glu	Asp	Gln	Lys 405	Arg	Asp	Thr	Ile	Asn 410	Leu	Leu	Asp	Gln	Arg 415	Glu

Lys Arg Asn His Thr Leu 420

```
<211> 14
  <212> PRT
  <213> Homo sapiens
  <400> 348
  Ser Leu His Arg Phe Val Leu Ser Gln Ala Lys Asp Glu Leu
                                       10
  <210> 349
 <211> 19
 <212> PRT
 <213> Homo sapiens
 <400> 349
 Phe Ile Lys Phe Phe Ala Pro Trp Cys Gly His Cys Lys Ala Leu Ala
                 5
                                     10
 Pro Thr Trp
 <210> 350
 <211> 19
 <212> PRT
 <213> Homo sapiens
 <400> 350
 Phe Ile Lys Phe Tyr Ala Pro Trp Cys Gly His Cys Lys Thr Leu Ala
 Pro Thr Trp
<210> 351
<211> 363
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (42)
<223> Xaa equals any of the naturally occurring L-amino acids
<400> 351
Arg Arg Gly Arg Gly Val Pro Gly Pro Arg Gly Arg Arg Arg Leu Trp
Ser Ala Ala Cys Gly His Cys Gln Arg Leu Gln Pro Thr Trp Asn Asp
Leu Gly Asp Lys Tyr Asn Ser Met Glu Xaa Ala Lys Val Tyr Val Ala
Lys Val Asp Cys Thr Ala His Ser Asp Val Cys Ser Ala Gln Gly Val
     50
                         55
Arg Gly Tyr Pro Thr Leu Lys Leu Phe Lys Pro Gly Gln Glu Ala Val
```

65	,				70					75	;				. 80
Lys	Туг	Gln	Gly	Pro 85		Asp	Phe	Gln	Thr 90		Glu	Asn	Trp	Met 95	Leu
Gln	Thr	Leu	Asn 100		Glu	Pro	Val	Thr 105		Glu	Pro	Glu	Val 110		Pro
Pro	Ser	Ala 115		Glu	Leu	Lys	Gln 120	_	Leu	Tyr	Glu	Leu 125	Ser	Ala	Ser
Asn	Phe 130		Leu	His	Val	Ala 135		Gly	Asp	His	Phe 140	Ile	Lys	Phe	Phe
Ala 145		Trp	Суѕ	Gly	His 150	Cys	Lys	Ala	Leu	Ala 155	Pro	Thr	Trp	Glu	Gln 160
Leu	Ala	Leu	Gly	Leu 165	Glu	His	Ser	Glu	Thr 170	Val	Lys	Ile	Gly	Lys 175	Val
Asp	Cys	Thr	Gln 180	His	Tyr	Glu	Leu	Cys 185	Ser	Gly	Asn	Gln	Val 190	Arg	Gly
Tyr	Pro	Thr 195	Leu	Leu	Trp	Phe	Arg 200	Asp	Gly	Lys	Lys	Val 205	Asp	Gln	Tyr
Lys	Gly 210	Lys	Arg	Asp	Leu	Glu 215	Ser	Leu	Arg	Glu	Tyr 220	Val	Glu	Ser	Gln
Leu 225	Gln	Arg	Thr	Glu	Thr 230	Gly	Ala	Thr	Glu	Thr 235	Val	Thr	Pro	Ser	Glu 240
Ala	Pro	Val	Leu	Ala 2 4 5	Ala	Glu	Pro	Glu	Ala 250	Asp	Lys	Gly	Thr	Val 255	Leu
Ala	Leu	Thr	Glu 260	Asn	Asn	Phe	Asp	Asp 265	Thr	Ile	Ala	Glu	Gly 270	Ile	Thr
Phe	Ile	Lys 275	Phe	Tyr	Ala	Pro	Trp 280	Cys	Gly	His	Cys	Lys 285	Thr	Leu	Ala
Pro	Thr 290	Trp	Glu	Glu	Leu	Ser 295	Lys	Lys	Glu	Phe	Pro 300	Gly	Leu	Ala	Gly
Val 305	Lys	Ile	Ala	Glu	Val 310	Asp	Суѕ	Thr	Ala	Glu 315	Arg	Asn	Ile	Cys	Ser 320
Lys	Tyr	Ser	Val	Arg 325	Gly	Tyr	Pro	Thr	Leu 330	Leu	Leu	Phe	Arg	Gly 335	
Lys	Lys	Val	Ser 340	Glu	His	Ser		Gly 345	Arg	Asp	Leu		Ser 350	Leu	His
Arg	Phe	Val 355	Leu	Ser	Gln	Ala	Lys 360	Asp	Glu	Leu					

<210> 352 <211> 93 <212> PRT <213> Homo sapiens

<400> 352

Met Arg Pro Gln Gly Pro Ala Ala Ser Pro Gln Arg Leu Arg Gly Leu

1 5 10 15

Leu Leu Leu Leu Gln Leu Pro Ala Pro Ser Ser Ala Ser Glu
20 25 30

Ile Pro Lys Gly Lys Gln Lys Ala His Ser Gly Arg Gly Arg Trp Trp 35 40 45

Thr Cys Ile Met Glu Cys Ala Tyr Lys Gly Gln Gln Glu Cys Leu Val

Glu Thr Gly Ala Leu Gly Pro Met Ala Phe Arg Val His Leu Gly Ser 65 70 75

Gln Val Gly Met Asp Ser Lys Glu Lys Arg Gly Asn Val

<210> 353

<211> 273

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (210)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 353

Glu Thr Arg Val Lys Thr Ser Leu Glu Leu Leu Arg Thr Gln Leu Glu

1 5 10 15

Pro Thr Gly Thr Val Gly Asn Thr Ile Met Thr Ser Gln Pro Val Pro
20 25 30

Asn Glu Thr Ile Ile Val Leu Pro Ser Asn Val Ile Asn Phe Ser Gln 35 40 45

Ala Glu Lys Pro Glu Pro Thr Asn Gln Gly Gln Asp Ser Leu Lys Lys
50 55 60

His Leu His Ala Glu Ile Lys Val Ile Gly Thr Ile Gln Ile Leu Cys
65 70 75 80

Gly Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe $85 \hspace{1cm} 90 \hspace{1cm} 95$

Ser Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr 100 105 110

Pro Phe Ile Gly Pro Phe Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile
115 120 125

Ala Thr Glu Lys Arg Leu Thr Lys Leu Leu Val His Ser Ser Leu Val

	130					133					140				
Gly 145	Ser	Ile	Leu	Ser	Ala 150	Leu	Ser	Ala	Leu	Val 155		Phe	lle	Ile	Leu 160
Ser	Val	Lys	Gln	Ala 165	Thr	Leu	Asn	Pro	Ala 170		Leu	Gln	Cys	Glu 175	
Asp	Lys	Asn	Asn 180	Ile	Pro	Thr	Arg	Ser 185	_	· Val	Ser	Tyr	Phe 190	_	His
Asp	Ser	Leu 195	Tyr	Thr	Thr	Asp	Cys 200	_	Thr	Ala	Lys	Ala 205		Leu	Ala
Gly	Xaa 210	Leu	Ser	Leu	Met	Leu 215		Cys	Thr	Leu	Leu 220	Glu	Phe	Cys	Leu
Ala 225	Val	Leu	Thr	Ala	Val 230	Leu	Arg	Trp	Lys	Gln 235		Tyr	Ser	Asp	Phe 240
Pro	Gly	Ser	Val	Leu 245	Phe	Leu	Pro	His	Ser 250		Ile	Gly	Asn	Ser 255	Gly
Met	Ser	Ser	Lys 260	Met	Thr	His	Asp	Суs 265	Gly	Tyr	Glu	Glu	Leu 270	Leu	Thr
Ser															
<211 <212)> 35 .> 19 !> PF !> Ho	92	sapie	ens											
<222	.> SI !> (1	.29)	r uals	any	of	the	nati	ural]	ly o	ccuri	ring	L-ar	mino	ació	ls
<400	> 35	. 4													
			Leu	Ser 5	Leu	Gly	Ile	Ile	Leu 10	Ala	Ser	Ala	Ser	Phe 15	Ser
Pro	Asn	Phe	Thr 20	Gln	Val	Thr	Ser	Thr 25	Leu	Leu	Asn	Ser	Ala 30	Tyr	Pro
Phe	Ile	Gly 35	Pro	Phe	Phe	Phe	Ile 40	Ile	Ser	Gly	Ser	Leu 45	Ser	Ile	Ala
Thr	Glu 50	Lys	Arg	Leu	Thr	Lys 55	Leu	Leu	Val	His	Ser 60	Ser	Leu	Val	Gly
Ser 65	Ile	Leu	Ser	Ala	Leu 70	Ser	Ala	Leu	Val	Gly 75	Phe	Ile	Ile	Leu	Ser 80
/al	Lys	Gln	Ala	Thr 85	Leu	Asn	Pro	Ala	Ser 90	Leu	Gln	Суѕ	Glu	Leu 95	Asp

- Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe Tyr His Asp 100 105 110
- Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala Ser Leu Ala Gly
 115 120 125
- Xaa Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Glu Phe Cys Leu Ala 130 135 140
- Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe Pro 145 155 160
- Gly Ser Val Leu Phe Leu Pro His Ser Tyr Ile Gly Asn Ser Gly Met
- Ser Ser Lys Met Thr His Asp Cys Gly Tyr Glu Glu Leu Leu Thr Ser 180 185 190
- <210> 355
- <211> 204
- <212> PRT
- <213> Homo sapiens
- <220>
- <221> SITE
- <222> (119)
- <223> Xaa equals any of the naturally occurring L-amino acids
- <400> 355
- Gly Ala Ser Cys Glu Gly Gly Gly Ala Ala Ala Arg Ala Ala Leu Gly
 1 5 10 15
- Val His Arg Ser Gln Lys Ala Leu Leu Val Phe Arg Arg Thr Leu Ser
- Asn Leu Leu Tyr Met Pro Leu Leu Arg Gly Leu Leu Trp Leu Gln Val 35 40 45
- Leu Cys Ala Gly Pro Leu His Thr Glu Ala Val Val Leu Leu Val Pro 50 55 60
- Ser Asp Asp Gly Arg Ala Phe Leu Leu Arg Ser Arg Leu Leu His Pro 65 70 75 80
- Glu Ala His Val Pro Pro Ala Ala Asp Arg Gly Ala Ser Leu Gln Cys
 85 90 95
- Val Leu His Gln Ala Ala Pro Lys Ser Arg Pro Arg Ser Pro Ala Ala 100 105 110
- Gly Ala Ala Leu Leu His Xaa Pro Arg Arg Thr Gly Asp Glu Pro Cys
 115 120 125
- Arg Glu Phe His Gly Asn Gly Phe Pro Gly Pro Thr Gln Leu Thr Pro
 130 135 140

Gly Glu Cys Gly Leu Pro Ala Pro Ser Ser Leu Leu Gln His Ala Ser 155

Ala Pro Val Arg Thr Gly Ser Glu Gly Gln Val Val Gly Cys Pro Arg

Ala Arg Gly Glu Thr Gly Glu Gly Leu Ser Leu Ala Phe Leu Ser Ser 180 185

Leu Met Phe Thr Ser Arg Asn Gly Leu Val Gly Cys

<210> 356

<211> 72

<212> PRT

<213> Homo sapiens

<400> 356

Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val

Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val

Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Thr Trp Lys

Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly Thr Cys Ser Ala

Lys Cys Thr Thr Arg Cys Trp Leu

<210> 357

<211> 115

<212> PRT

<213> Homo sapiens

<400> 357

Leu Lys Arg Ala Pro Pro Gly Pro Ala Leu Ala Lys Gly Leu Leu Gln

Pro Ser Ser Thr Phe Gln Ala Leu Glu Thr Asn Ile Gly Asp Gln Val 25 30

Arg Arg His Ser Thr Ala Val Val Ile Arg Glu Met Thr Ser Tyr Ile

Leu Ile Ser Phe Val Leu Leu Ile Gly Val Gly Cys Ile Glu Lys Asp

Gln Ser Cys Pro Val Phe Gly Gly Arg Lys Arg Leu His Leu Leu Phe

Val Gly Gly Gln Leu Arg Gln Val Arg Met Leu Arg Gly Glu Leu Ser

```
Cys Ala Cys Tyr Arg Pro His Val Gln Ala Leu Gln Leu Gly Gly Cys
  Thr Cys Phe
          115
  <210> 358
  <211> 88
  <212> PRT
  <213> Homo sapiens
  <400> 358
 Val Ile Lys Leu Ile Cys Pro Ala Ala Phe Pro Val Tyr Phe Gln Asp
                                       10
 Met Ala Arg Gly Cys Val Cys Ser Leu Cys Ala Ser Val Cys Ile Phe
                                   25
 Leu Ser Ser Leu Phe Pro Leu Leu Pro Ser Val His Ser Val Asn Ile
                               40
 Ile Ser Cys Leu Leu Ser Lys Cys Phe Glu Gly Leu Glu Leu Met
                         · 55
 Cys Glu His Leu Tyr Gln Leu Ser Gln Leu His Val Leu His His Ile
                      70
 Phe Ser Tyr Leu Leu Cys Thr Pro
                  85
 <210> 359
 <211> 716
 <212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> (2)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (373)
<223> Xaa equals any of the naturally occurring L-amino acids
<220>
<221> SITE
<222> (705)
<223> Xaa equals any of the naturally occurring L-amino acids
<400> 359
Tyr Xaa Ile Pro Gly Ser Thr His Ala Ser Gly Arg Gln Arg Gly Ser
```

Gly Arg Gly Glu Asp Asp Ser Gly Pro Pro Pro Ser Thr Val Ile Asn

25

20

Gln	Asn	Glu 35		Phe	. Ala	Asn	Ile 40		e Phe	E Lys	Pro	Thr 45		. Val	Ğl
Gln	Ala 50		Ile	Ala	Gln	Asn 55	Gly	Ile	. Lev	ı Gly	Asp 60		: Ile	: Ile	Ar
Tyr 65	Asp	Val	Asn	Arg	Glu 70		Ser	Ile	Gly	Asp 75		Gln	Val	. Leu	. As:
Gly	Tyr	Phe	Val	His 85	-	Phe	Ala	Pro	Lys 90	_	Leu	Pro	Pro	Leu 95	
Lys	Asn	Val	Val 100		Val	Leu	Asp	Ser 105		Ala	Ser	Met	Val 110	_	Th:
Lys	Leu	Arg 115		Thr	Lys	Asp	Ala 120	Leu	Phe	Thr	Ile	Leu 125		Asp	Lei
Arg	Pro 130	Gln	Asp	Arg	Phe	Ser 135	Ile	Ile	Gly	Phe	Ser 140	Asn	Arg	Ile	Lys
Val 145	Trp	Lys	Asp	His	Leu 150	Ile	Ser	Val	Thr	Pro 155	Asp	Ser	Ile	Arg	Ası 160
				165			Met		170					175	
			180				Ile	185					190		
		195					Ser 200					205			
	210					215	Glu				220	_			
225					230		Gly			235					240
				245			Arg		250					255	
			260				Val	265					270		
		275					Glu 280					285			
	290					295	Ser				300				
305					310		Gly			315					320
Leu	Val	Asp	Arg	Lys	Leu	Asp	His	Leu	His	Val	Glu	Val	Thr	Ala 335	Ser

Asn Ser Lys Lys Phe Ile Ile Leu Lys Thr Asp Val Pro Val Arg Pro

	•		
3	40	345	350
Gln Lys Ala G	ly Lys Asp Val	Thr Gly Ser Pro Ar	g Pro Gly Gly Asp 365
Gly Glu Gly As 370	sp Xaa Asn His 375	Ile Glu Arg Leu Tr 38	p Ser Tyr Leu Thr 0
Thr Lys Glu Le 385	eu Leu Ser Ser 390	Trp Leu Gln Ser As	p Asp Glu Pro Glu 400
Lys Glu Arg Le	u Arg Gln Arg 405	Ala Gln Ala Leu Ala 410	a Val Ser Tyr Arg 415
Phe Leu Thr Pr 42	o Phe Thr Seri	Met Lys Leu Arg Gly 425	/ Pro Val Pro Arg
Met Asp Gly Le 435	u Glu Glu Ala i	His Gly Met Ser Ala 440	Ala Met Gly Pro
Glu Pro Val Val 450	l Gln Ser Val A 455	Arg Gly Ala Gly Thr 460	Gln Pro Gly Pro
Leu Leu Lys Lys 465	Pro Tyr Gln F	Pro Arg Ile Lys Ile 475	Ser Lys Thr Ser
Val Asp Gly Asp	Pro His Phe V 485	al Val Asp Phe Pro 490	
Thr Val Cys Phe 500	Asn Ile Asp G	ly Gln Pro Gly Asp 505	• •
Val Ser Asp His 515	Arg Asp Ser G	ly Val Thr Val Asn 20	
Gly Ala Pro Ala 530	Pro Pro Asn Gi 535	ly His Lys Lys Gln	Arg Thr Tyr Leu
Arg Thr Ile Thr 545	Ile Leu Ile As 550	on Lys Pro Glu Arg 555	Ser Tyr Leu Glu 560
Ile Thr Pro Ser	Arg Val Ile Le 565	eu Asp Gly Gly Asp 7	Arg Leu Val Leu 575
Pro Cys Asn Gln 580	Ser Val Val Va	l Gly Ser Trp Gly 1 585	Leu Glu Val Ser 590
Val Ser Ala Asn . 595	Ala Asn Val Th 60	r Val Thr Ile Gln G	Gly Ser Ile Ala
Phe Val Ile Leu : 610	Ile His Leu Ty: 615	r Lys Lys Pro Ala F 620	
His His Leu Gly I 625	Phe Tyr Ile Ala 630	a Asn Ser Glu Gly L 635	eu Ser Ser Asn 640
Cys His Gly Leu I	Leu Gly Gln Phe 145	e Leu Asn Gln Asp A 650	

245

Glu Asp Pro Ala Gly Pro Ser Gln Asn Leu Thr His Pro Leu Leu Leu 660 665 670

Gln Val Gly Glu Gly Pro Glu Ala Val Leu Thr Val Lys Gly His Gln 675 680 685

Val Pro Val Val Trp Lys Gln Arg Lys Ile Tyr Asn Gly Glu Glu Gln 690 695 700

Xaa Asp Cys Trp Phe Ala Arg Asn Met Pro Pro Asn 705 710 715

<210> 360

<211> 387

<212> PRT

<213> Homo sapiens

<400> 360

Pro Arg Val Arg Ser Ile Lys Val Thr Glu Leu Lys Gly Leu Ala Asn

1 5 10 15

His Val Val Gly Ser Val Ser Cys Glu Thr Lys Asp Leu Phe Ala 20 25 30

Ala Leu Pro Gln Val Val Ala Val Asp Ile Asn Asp Leu Gly Thr Ile 35 40 45

Lys Leu Ser Leu Glu Val Thr Trp Ser Pro Phe Asp Lys Asp Asp Gln 50 55 60

Pro Ser Ala Ala Ser Ser Val Asn Lys Ala Ser Thr Val Thr Lys Arg 65 70 75 80

Phe Ser Thr Tyr Ser Gln Ser Pro Pro Asp Thr Pro Ser Leu Arg Glu 85 90 95

Gln Ala Phe Tyr Asn Met Leu Arg Arg Gln Glu Glu Leu Glu Asn Gly 100 105 110

Thr Ala Trp Ser Leu Ser Ser Glu Ser Ser Asp Asp Ser Ser Ser Pro 115 120 125

Gln Leu Ser Gly Thr Ala Arg His Ser Pro Ala Pro Arg Pro Leu Val 130 135 140

Gln Gln Pro Glu Pro Leu Pro Ile Gln Val Ala Phe Arg Arg Pro Glu 145 150 155 160

Thr Pro Ser Ser Gly Pro Leu Asp Glu Glu Gly Ala Val Ala Pro Val 165 170 175

Leu Ala Asn Gly His Ala Pro Tyr Ser Arg Thr Leu Ser His Ile Ser 180 185 190

Glu Ala Ser Val Asn Ala Ala Leu Ala Glu Ala Ser Val Glu Ala Val 195 200 205

Gly Pro Lys Ser Leu Ser Trp Gly Pro Ser Pro Pro Thr His Pro Ala

210 215 Pro Thr His Gly Lys His Pro Ser Pro Val Pro Pro Ala Leu Asp Pro 230 Gly His Ser Ala Thr Ser Ser Thr Leu Gly Thr Thr Gly Ser Val Pro 250 Thr Ser Thr Asp Pro Ala Pro Ser Ala His Leu Asp Ser Val His Lys 260 265 Ser Thr Asp Ser Gly Pro Ser Glu Leu Pro Gly Pro Thr His Thr Thr 280 Thr Gly Ser Thr Tyr Ser Ala Ile Thr Thr Thr His Ser Ala Pro Ser Pro Leu Thr His Thr Thr Gly Ser Thr His Lys Pro Ile Ile Ser 310 Thr Leu Thr Thr Gly Pro Thr Leu Asn Ile Ile Gly Pro Val Gln 325 330 Thr Thr Thr Ser Pro Thr His Thr Met Pro Ser Pro Ser Ser His Ser Asn Ser Pro Gln Tyr Val Asp Phe Cys Ser Ser Val Cys Asp Asn Ile 355 365 Phe Val His Tyr Val Ile Gly Ile Phe Phe His Thr Leu Tyr Ser Ser 375 Lys Thr Leu 385 <210> 361 <211> 260 <212> PRT <213> Homo sapiens <400> 361 Pro Arg Val Arg Ser Ile Lys Val Thr Glu Leu Lys Gly Leu Ala Asn His Val Val Gly Ser Val Ser Cys Glu Thr Lys Asp Leu Phe Ala 20 25 Ala Leu Pro Gln Val Val Ala Val Asp Ile Asn Asp Leu Gly Thr Ile 40 Lys Leu Ser Leu Glu Val Thr Trp Ser Pro Phe Asp Lys Asp Asp Gln 55 Pro Ser Ala Ala Ser Ser Val Asn Lys Ala Ser Thr Val Thr Lys Arg Phe Ser Thr Tyr Ser Gln Ser Pro Pro Asp Thr Pro Ser Leu Arg Glu

85

Gln Ala	Phe	Туг 100	Asn	Met	Leu	Arg	Arg 105	Gln	Glu	Glu	Leu	Glu 110	Asn	Gly
Thr Ala	Trp 115	Ser	Leu	Ser	Ser	Glu 120	Ser	Ser	Asp	Asp	Ser 125	Ser	Ser	Pro
Gln Leu 130	Ser	Gly	Thr	Ala	Arg 135	His	Ser	Pro	Ala	Pro 140	Arg	Pro	Leu	Val
Gln Gln 145	Pro	Glu	Pro	Leu 150	Pro	Ile	Gln	Val	Ala 155	Phe	Arg	Arg	Pro	Glu 160
Thr Pro	Ser	Ser	Gly 165	Pro	Leu	Asp	Glụ	Glu 170	Gly	Ala	Val	Ala	Pro 175	Val
·Leu Ala	Asn	Gly 180	His	Ala	Pro	Tyr	Ser 185	Arg	Thr	Leu	Ser	His 190	Ile	Ser
Glu Ala	Ser 195	Val	Asn	Ala	Ala	Leu 200	Ala	Glu	Ala	Ser	Val 205	Glu	Ala	Val
Gly Pro 210		Ser	Leu	Ser	Trp 215	Gly	Pro	Ser	Pro	Pro 220	Thr	His	Pro	Ala
Pro Thr 225	His	Gly	Lys	His 230	Pro	Ser	Pro	Val	Pro 235	Pro	Ala	Leu	Asp	Pro 240
Gly His	Ser	Ala	Thr 245	Ser	Ser	Thr	Leu	Gly 250	Thr	Thr	Gly	Ser	Val 255	Pro
Thr Ser	Thr	Asp 260												
<210> 3 <211> 1 <212> P <213> H	55 RT .	sapie	ens											•
<400> 3 Tyr Gly		Glu	Lys 5	Thr	Thr	Glu	Gly	Gly 10	Arg	Arg	Arg	Arg	Arg 15	Arg
Met Glu -	Ala	Val 20	Val	Phe	Val	Phe	Ser 25	Leu	Leu	Asp	Cys	Cys 30	Ala	Leu
Ile Phe	Leu 35	Ser	Val	Tyr	Phe	11e 40	Ile	Thr	Leu	Ser	Asp 45	Leu	Glu	Cys
Asp Tyr 50		Asn	Ala	Arg	Ser 55	Cys	Суѕ	Ser	Lys	Leu 60	Asn	Lys	Trp	Val

Ile Pro Glu Leu Ile Gly His Thr Ile Val Thr Val Leu Leu Met 65 70 75 80

Ser Leu His Trp Phe Ile Phe Leu Leu Asn Leu Pro Val Ala Thr Trp

90

Asn Ile Tyr Arg Tyr Ile Met Val Pro Ser Gly Asn Met Gly Val Phe 100 105 110

Asp Pro Thr Glu Ile His Asn Arg Gly Gln Leu Lys Ser His Met Lys 115 120 125

Glu Ala Met Ile Lys Leu Gly Phe His Leu Leu Cys Phe Phe Met Tyr 130 135 140

Leu Tyr Ser Met Ile Leu Ala Leu Ile Asn Asp 145 150 155

<210> 363

<211> 70

<212> PRT

<213> Homo sapiens

<400> 363

Ala Arg Ala Pro Ala Pro Ser Leu Pro Pro Leu Pro Ser Pro Ala Pro 1 5 10 15

Ala Leu Ala Pro Ala His Ser Leu Leu Gly Leu Leu Gly Arg Met 20 25 30

Ser Gly Ser Ser Leu Pro Ser Ala Leu Ala Leu Ser Leu Leu Leu Val

Ser Gly Ser Leu Leu Pro Gly Pro Gly Ala Ala Gln Asn Val Arg Val 50 55 60

Gln Ser Gly Gln Asp Gln

<210> 364

<211> 56

<212> PRT

<213> Homo sapiens

<400> 364

Gly Thr Ser Lys Asp Cys Val Leu Tyr Ala Phe Leu Asp Pro Gly Met

Ala Val Pro Leu Phe Leu Tyr Ile Phe Thr Leu Leu Pro Leu Leu Pro 20 25 30

Phe Leu Leu Ser Leu Cys Phe Ser Pro Leu Thr Val Lys Arg Ser Ser 35 40 45

Ser Ser Glu Ser Lys Ser Ser Leu
50 55

Applicant's or agent's file	D744400T	International application	
reference number	PZ031PCT	1	Unassigned
TOTOT CITE TO THE COLOR			

(PCT Rule 13bis) REC'D 1 8 AUG 1999 A. The indications made below relate to the microorganism referred to in the description **WIPO** PCT 260 on page , line B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America Date of deposit Accession Number July 27, 1998 203069 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States) In respect to those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC). E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit") For receiving Office use only For International Bureau use only This sheet was received with the international application This sheet was received by the International Bureau on: Authorized officer Authorized officer PCT Gernsumal Division

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

Applicant's or agent's file reference number	PZ031PCT	International application	Unassigned
			 Unassigned

(PCT Rule 13bis)

A	The indication	s made helow t		
	on page	243	microorganism re , line	eferred to in the description N/A
B.	IDENTIFICA'	TIONOFDEPOSIT		Further deposits are identified on an additional sheet
Na	me of depositary	institution American T	ype Culture Co	ollection
_				•
		ary institution (including	postal code and co	ountry)
Ma	801 University Inassas, Virgir	nia 20110-2209		•
Un	ited States of	America		
Dat	e of deposit			
υaι	e or deposit	June 11, 1998		Accession Number 209965
 C.	ADDITIONAL	L INDICATIONS (leave	blank if not applied	
				This information is continued on an additional sheet
	•			
_				
). 1	DESIGNATED	STATES FOR WHIC	TH INDICATIO	ONS ARE MADE (if the indications are not for all designated States)
uro	pe			
	oraniani wiii	ue illade avallable u	ntil the bublicat	Patent is sought a sample of the deposited tion of the mention of the grant of the European patent
		"""" BUUILAUUII NA	s neen remsec	d or withdrawn or is deemed to be withdrawn, only by the person requesting the sample (Rule 28(4) EPC).
		The to all expert	norminated by	the person requesting the sample (Hule 28(4) EPC).
. s	EPARATEFU	RNISHING OF INDIC	CATIONS (leave	blankifnot applicable)
he i				mal Bureau later (specify the general nature of the indications e.g., "Accession
				·
				·
_	Forre	ceiving Office use only		
(1		eived with the internation	al application	For International Bureau use only This sheet was received by the International Bureau on:
_)	— The international Bureau on:
tho	rizedofficer	- Jan - Comp	کننځ	. Authorized officer
		PCT memorianal DI	vision	1
_				<u></u>

Form PCT/RO/134 (July 1992)

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

Applicant's or agent's file	· · · · · · · · · · · · · · · · · · ·	. International continuing	
reference number	PZ031PCT	International application	Unassigned
TOTO TO			

(PCT Rule 13bis)

A The	a indiantions	manda halassas tasas at		
	page	249	microorganism refe , line	erred to in the description N/A
B. IDI	ENTIFICAT	ΠΟΝΟFDEPOSIT		Further deposits are identified on an additional sheet
			vno Cultura Call	
IVAILLEO	o depositary	institution American T	ype Culture Coll	ection
				·
Address	s of deposit	ary institution (including	postal code and cou	niry)
10801	University	Boulevard		
	ssas, Virgir States of	nia 20110-2209	-	
Omico	Jiales UI	America		
Data of	danni.			
Date of	aeposii	June 26, 1998		Accession Number
				203027
C. AD	DITIONAL	L INDICATIONS (leav	e blank if not applicat	ole) This information is continued on an additional sheet
		•		
D. DES	SIGNATEI	STATES FOR WHI	CH INDICATIO	NS ARE MADE (if the indications are not for all designated States)
urope			·	
n respe	ect to those	designations in which	h a European F	Patent is sought a sample of the deposited
or until t	the date or	i de made avallable to n which application ha	intil the publicati as been refused	ion of the mention of the grant of the European patent or withdrawn or is deemed to be withdrawn, only by
he issu	e of such a	a sample to an experi	t nominated by t	he person requesting the sample (Rule 28(4) EPC).
		-		
E. SEP	ARATEF	URNISHING OF INDI	CATIONS (leave l	blank if not applicable)
The indic	cations lister of Deposit")	d below will be submitte	d to the Internation	nal Bureau later (specify the general nature of the indications e.g., "Accession
	, Deposii ,			
		•		
	Forr	eceiving Office use only		For International Bureau use only
This	s sheet was re	ceived with the internatio	nal application	This sheet was received by the International Bureau on:
•	•	! .	1	
Authorize	ed officer	114	m	Authorized officer
	· PA	vette E. Storms Of International Divi	cion I	[
			01.	

Form PCT/RO/134 (July 1992)

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

Applicant's or agent's file reference number	PZ031PCT	International application	Unassigned
			J

(PCT Rule 13bis)

A. The indic	ations made below relat	e to the microorganism r	referred to in the description
on page	253	, line	N/A
B. IDENTII	FICATIONOFDEPOS	err	Further deposits one id- vic
			Further deposits are identified on an additional sheet
rvanie or depo	sitary institution Amer	ican Type Culture C	ollection
			·
Address of de	epositary institution (in	cluding postal code and c	country)
	ersity Boulevard		
Manassas, '	Virginia 20110-220:	9	
United State	es of America		
Date of deposi	I .		Accession Number
	July 27, 19	98	203071
C ADDITIO	ONAL INDICATION	IS (leave blank if not applic	orble) This is formation in
			cable) This information is continued on an additional sheet
		•	
	-		N.
DESIGNA	TED STATES FOR	WHICH INDICATION	ONS ARE MADE (if the indications are not for all designated States)
rope			ONO MICE INTERDECT THE trutcusions are not for all designated States)
respect to t	hose designations i	n which a European	Patent is sought a sample of the deposited
ici ooi yaa iisi	ii wiii oe made avail	able until the publics	Stinn of the mention of the growt of the E
C 010 02	re on which applicat	ion nas been reitise	Of Wilhdrawn or is deemed to be withdrawn and t
- 10020 01 01	son a sample to an e	expert nominated by	the person requesting the sample (Rule 28(4) EPC).
			
		INDICATIONS (leave	
ne indications unber of Depos	listed below will be su	bmitted to the Internati	onal Bureau later (specify the general nature of the indications e.g., "Accession
macr of Depos	<i></i> ,		•
			·
	Enganasista - OFF		
This sheer	For receiving Office use was received with the inte	•	For International Bureau use only
۲عسوور <i>ب</i>	as received with the inte	mational application	This sheet was received by the International Bureau on:
		X	
thorized office	The symm	Jun	Authorized officer
	Bot lawaries	al Division]]
	_		7 I

Form PCT/RO/134 (July 1992)

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

			•
Applicant's or agent's file reference number	PZ031PCT	International application	

(PCT Rule 13bis)

A The indigesions made below	
A. The indications made below relate to the microorganism re on page	oferred to in the description N/A
B. IDENTIFICATIONOF DEPOSIT	Further deposits are identified on an additional sheet
Name of depositary institution American Type Culture Co	
and the control of th	one chorr
Address of depositary institution (including postal code and co	
	nuntry)
10801 University Boulevard Manassas, Virginia 20110-2209	
United States of America	
Date of deposit	Accession Number
July 27, 1998	203070
C ADDITIONAL INDICATIONS (leave blank if not applice	able) This information is continued on an additional sheet
D. DESIGNATED STATES FOR WHICH INDICATION	ONS ARE MADE (if the indications are not for all designated States)
шгоре	The state of the s
respect to those designations in which a European	Patent is sought a sample of the deposited
icroorganism will be made available until the publica	tion of the mention of the grant of the European patent d or withdrawn or is deemed to be withdrawn, only by
e issue of such a sample to an expert nominated by	the person requesting the sample (Rule 28(4) EPC).
	, , , , , , , , , , , , , , , , , , , ,
. SEPARATE FURNISHING OF INDICATIONS (leave	e blank if not applicable)
he indications listed below will be submitted to the Internation	onal Bureau later (specify the general nature of the indications e.g., "Accession
umber of Deposii")	,
	4
For receiving Office use only	For International Bureau use only
This sheet was received with the international application	This sheet was received by the International Bureau on:
	,
uthorized officer Simms	Authorized officer
PCY International Division	

Form PCT/RO/134 (July 1992)

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/17130

A. CLASSIFICATION OF SUBJECT MATTER IPC(6) :Please See Extra Sheet.							
US CL : Please See Extra Sheet.							
According to International Patent Classification (IPC) or to b	oth national classification and IPC						
B. FIELDS SEARCHED							
Minimum documentation searched (classification system follo							
U.S. : 536/23.1, 23.5; 435/69.1, 320.1, 252.3, 325, 6, 7.	1; 530/350, 300, 387.1; 514/2						
Documentation searched other than minimum documentation to	o the extent that such documents are included in the fields searched						
Electronic data base consulted during the international search	(name of data base and, where practicable, search terms used)						
APS, DIALOG - Biotech Files	, , , , , , , , , , , , , , , , , , , ,						
C. DOCUMENTS CONSIDERED TO BE RELEVANT							
The state of the s	The state of the s						
A JACOBS, K. A. et al. A Genetic Encoding Secreted Proteins. Gene.	Selection For Isolating cDNAs 1-23						
296, see entire document.	, , , , , , , , , , , , , , , , , , ,						
	Ĺ						
·							
·	,						
	·						
Further documents are listed in the continuation of Box	C. See patent family annex.						
Special estegories of cited documents:	"T" later document published after the international Gling date.						
A* document defining the general state of the art which is not considered to be of particular relevance	date and not in conflict with the application but cited to understand the principle or theory underlying the invention						
B* earlier document published on or after the international filling date L* document which may throw doubt an arceits shirt() and the	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step						
L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	witer the document is taken glone						
O* document referring to an oral disclosure, use, exhibition or other means	'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art						
P document published prior to the international filing date but later than the priority date claimed	*&* document member of the same patent family						
Date of the actual completion of the international search	Date of mailing of the international search report						
05 OCTOBER 1999	21 OCT 1999						
lame and mailing address of the ISA/US Commissioner of Patents and Trademarks	Authorized officer						
Box PCT Washington, D.C. 20231	ELIZABETH C. KEMMERER						
acsimile No. (703) 305-3230	Telephone No. (703) 308-0196						

INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/17130

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
Claims Nos.: 1-23 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: Please See Extra Sheet.
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest The additional search fees were accompanied by the applicant's protest.
No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/17130

A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

C12N 1/21, 5/10, 15/11, 15/12, 15/63; A61K 38/16, 38/17; C07K 14/00, 14/435, 16/00; G01N 33/50

A. CLASSIFICATION OF SUBJECT MATTER: US CL:

536/23.1, 23.5; 435/69.1, 320.1, 252.3, 325, 6, 7.1; 530/350, 300, 387.1; 514/2

BOX I. OBSERVATIONS WHERE CLAIMS WERE FOUND UNSEARCHABLE 2. Where no meaningful search could be carried out, specifically:

All of the claims were unsearchable to the extent that they require reference to sequences from the sequence listing or an ATCC deposit. However, the specific sequence and deposit numbers were replaced in the claims with generic designators X, Y and Z. Therefore, no meaningful search of the sequences or deposits per se can be carried out by this Authority. The subject matter of the claims has been searched only to the extent possible with reference to the balance of the description.